DOCUMENT RESUME

ED 093 962 TM 003 780

AUTHOR	Benedict, Larry
TITLE	Practical Guide for (Educational) Evaluation.
INSTITUTION	Capitol Region Education Council, West Hartford, Conn.
PUB DATE	74
NOTE	91p.
EDRS PRICE	MF-$0.75 HC-$4.20 PLUS POSTAGE
DESCRIPTORS	*Administrator Guides; *Decision Making; Educational Improvement; Educational Needs; *Evaluation; Evaluation Methods; *Guides

ABSTRACT

This booklet deals with the practical steps of educational evaluation: who should negotiate the contract? Who initiates evaluation? What are "goals process" and "parts process" and how are they matched? What are the steps in putting the process of evaluation into operation? What are the criteria for assessing observational techniques? What data need to be collected? Once a decision-maker has a report of evaluation, what will he do with it? Suppose a school district has limited resources, what can the decision-maker do? A glossary is provided at the end of the booklet so that the reader will know what the writer means by some words or terms used. This guide will be followed by booklets addressed to specific audiences: e.g., members of the boards of education, administrators, teachers and parents. (BB)

ED 093962

TM 003 280

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

BEST COPY AVAILABLE

PROJECT EVALUATION
CAPITOL REGION EDUCATION COUNCIL

PRACTICAL GUIDE FOR EVALUATION

PREPARED BY: DR. LARRY BENEDICT
UNIVERSITY OF MASSACHUSETTS

TABLE OF CONTENTS

		Page
I.	An Introduction to Educational Evaluation	1
	Review	3
	Some Basic Concepts of Evaluation	4
	Decision-Maker and Decision Making	6
II.	The First Step in Evaluation	10
	Negotiation of the Contract: Initiation of the Evaluation	10
	Review	15
	Preparation of the Evaluation Contract	16
III.	A Goals Process	17
	Review	22
IV.	A Parts Process	23
	Review	32
V.	A Matching Process for Goals and Parts	33
	Review	35
VI.	An Operationalization Process	36
	Review	50
VII.	Measurement for Evaluation	51
	Criteria to Assess Observational Techniques	54
	Review	56
VIII.	Data Collection	57
	Review	63
IX.	Having Evaluation Data Reported to the Decision-Maker	64
	Review	67
	What a Report Should Not Have	68
	Review	70
X.	Redesigning the Evaluation	71
	Review	74

TABLE OF CONTENTS
(cont'd)

Page

XI. Evaluation of Evaluation.....75

Review.....78

XII. When Resources for the Evaluation are Really Small, What do you do?.....79

XIII. Glossary of Terms.....81

XIV. References Used in the Text.....83

XV. Additional References Which Might Be Used as Resources.....85

BEST COPY AVAILABLE

Preface

This booklet has been prepared as a guide for decision-makers in education (Board members, administrators and teachers. . .) who may hire an evaluator to begin an evaluation. The user will find it a very helpful manual in delineating evaluation problems.

After dealing with some basic concepts of evaluation to clarify misunderstandings and misinformation, Dr. Benedict, who prepared this booklet, deals with the practical steps of evaluation: Who should negotiate the contract? Who initiates evaluation? What are "goals process" and "parts process" and how are they matched? What are the steps in putting the process of evaluation into operation? What are the criteria for assessing observational techniques? What data need to be collected? Once a decision-maker has a report of evaluation, what will he do with it? Suppose a school district has limited resources, what can the decision-maker do?

Dr. Benedict has tried, and I believe he has succeeded, to avoid some of the educational terminology that has "fuzzy" meaning. However, a glossary is provided at the end of the booklet so that the reader will know what the writer means by some words or terms used.

This guide will be followed by booklets addressed to specific audiences: e.g., members of the boards of education, administrators, teachers and parents.

In introducing this guide, Project Evaluation, Capitol Region Education Council, considers it a step on the long road of evaluation. I hope that it will be widely used among decision-makers towards the betterment of the educational process.

Philip S. Saif, Director
Project Evaluation

BEST COPY AVAILABLE

I. AN INTRODUCTION TO EDUCATIONAL EVALUATION

The starting point in evaluation occurs well before the evaluation begins. That point should be when one asks, and answers, the question: "Why do I want to evaluate?" Unless this question is answered, an evaluation should not be undertaken because, in fact, maybe it is not evaluation that is needed or wanted, but something else.

Here are some typical reasons for wanting to have an evaluation:

(1) For public relations - so someone will like me, or fund me, etc.

(2) To find out what the students need.

(3) To make program or planning decisions.

(4) To provide systematic, ongoing information (data) as a basis for making decisions.

However, not all of these are evaluation, so a decision-maker would not (should not) hire an evaluator to do all of these. For example, evaluation is fundamentally different from a public relations job. PR brings to mind Madison Avenue, marketing, public image and so on. This is not to say that a PR man might not want to avail himself of some of the data an evaluation design would collect. This is to say, however, that the evaluation designer's job is not PR. If an enterprise wishes to sell itself to the public, it hires a PR expert, goes to an advertising agency or buys commercial time. If an enterprise desires objective, systematic feedback about the status of that enterprise, it hires an evaluator or evaluation designer.

It is important not to confuse the roles of PR and evaluation, for the methods, nature and goals of each are fundamentally different. A PR expert is in a much better position to do a much better job of promoting one's image or selling one's wares than is a person trained only in evaluation. Conversely, a PR man is not usually equipped or skilled in evaluation design. Basically, then, this simple rule of thumb should be remembered: if one wants a PR job, hire a PR man; if one wants an evaluation design, hire an evaluator.

The same can be said of Purpose #2. This purpose really demands a needs analysis expert, not a person skilled in evaluation. While the two may be similar, a needs analysis can be better done by someone trained in such procedures, rather than someone trained in evaluation.

Purpose #3 above is also not evaluation. Making program or planning decisions is decision making. If an enterprise wants to hire someone to make decisions for them, to improve their decision making, to insure that the enterprise makes "good" or "the right" decisions, then the enterprise should hire someone trained in decision making.

The fourth purpose is the one being agreed upon by more and more "evaluation experts". Evaluation has as its primary purpose the collection of data to be used as feedback to decision-makers in order to provide a basis for decision making, not to make decisions. It is more than assessing student achievement, more than measuring the percentage of achievement of an instructional objective. Rather, evaluation should be the collection of specific data about a given program or project which the decision-makers of that project want or that the enterprise deems important and which will be used by those decision-makers for decision making regarding the strengths and weaknesses of their particular enterprise.

Review: An Introduction to Educational Evaluation

(1) The first step before an evaluation is begun is to determine the purpose for conducting it.

(2) If your purpose is to have data for decision making, then you are in the same ball park as educational evaluation experts (Cronbach, Guba, Stufflebeam, Fortune, Hutchinson, Vorthen, Provus, and many others).

(3) If your purpose is not to collect data for your decision making needs, but some other purpose, seek an expert in that ball park.

Having come up with an answer to "Why do I want to evaluate? the next step is to consider some basic concepts of evaluation.

Some Basic Concepts of Evaluation

The term "evaluation" is an all-encompassing concept in education today. Many, many processes are termed "evaluation" when in fact they would probably be better termed something else. Some examples will show how fuzzy a concept "evaluation" is.

> The testing of products to describe their characteristics is called evaluation. Why not simply call it product testing?
>
> The accumulation of data about an institution's operation - its income, expenditures, costs per credit hour, faculty-student ratio, etc. is called evaluation. Why not simply call it institutional accounting?
>
> The measurement of pupils' knowledge at the beginning and end of a course is called evaluation, Why not simply call it achievement testing? (Pace, 1968, pp. 1-2)

These are a few examples which show some of the different things called evaluation. Yet each of these is not evaluation. Evaluation is different. The purpose of this section is to discuss what is and is not evaluation.

Traditionally, evaluation has been conceived of as the administering of a test, usually standardized, for the purpose of determining something, usually student achievement. Or secondly, evaluation has been traditionally conceived of as determining "how good" or "how bad" something stacks up to something else, i.e., Program A to Program B, or School A to School B.

This approach can be labeled the Traditional Model of Evaluation. It is usually implemented in the following manner: an outside expert (consultant) is hired to do an evaluation. He looks around for a few days to get a "feel" for the enterprise, selects a set of standardized tests that he thinks have something to do with the enterprise and administers them, both pre- and post-. The results, showing no significant differences, are written up in the form of a critical report. Finally, the "evaluator" collects his fee and possibly publishes the report.

The major purpose of this model seems to be the professional development of the evaluator. Thus another possible title for this model is the Evaluator Model, or the Evaluator-as-Expert Model.

This is not a legitimate function of, nor purpose for evaluation. Furthermore, it is not even a sound procedure for conducting an evaluation, e.g., simply pre- and post-testing. Although "evaluation" and "testing" have usually been used interchangeably in educational research, evaluation is more than just testing.

This conception - Evaluator-as-Expert-Model - of evaluation is both narrow and usually not very useful to the decision-makers for whom it is done. In terms of the decision-makers involved, these types of evaluations provide little if any useful data on which to make decisions regarding program strengths and weaknesses, redefinition and refining of program processes, etc. This in fact explains to some extent why so many seemingly excellent evaluations (excellent at least from the perspective of the evaluator) have been written, bound and put on the shelf, there to remain unopened and unread, their conclusions and recommendations being ignored, not acted upon. And the evaluator who conducted the study can't understand why such an "excellent" report is being ignored by the project's decision-makers. He fails to realize that the data which are not being used by the decision-makers must not be relevant to them and their needs, and that this factor is due in part to his own narrow conception of evaluation.

The function of evaluation must be to provide relevant data to some decision-makers with respect to some project, i.e., data they will use for decision making. This is, it will be noticed, a much more useful concept of evaluation than the pre-post test approach and administering of a standardized test at the end of the year.

Another traditional approach to evaluation has been to have a Board of Experts come into an enterprise to do the "evaluation". This is found in its highest form in the Accreditation Model, with which most school personnel are familiar. The Accreditation Team looks at the physical plant, number of chairs, number of books, etc. It doesn't really look at program outcomes. Such reports are usually very descriptive about very "physical"things.

Quality of learning seldom enters the picture. The real concerns of the enterprise's decision-makers are not the focus of such "evaluations".

However, moving away from these traditional concepts of evaluation, it is not only possible but essential to discuss a more effective and useful concept of evaluation. As Stufflebeam has written,

> Evaluation is a science of relating antecedent conditions and processes to outcomes and outcomes to objectives. Evaluation strives (1) to determine the extent to which objectives are achieved - to measure and define outcomes, and (2) to uncover the functional relationships between outcome and process variables - to explain outcomes. (1967a, p. 127)

This definition is not necessarily inconsistent with the pre-post test approach. However, it has to be taken in conjunction with another concept, namely, that of "decision-maker" and "decision making".

Decision Maker and Decision Making

This concept is a relatively new one in the history of educational evaluation. In 1963 Cronbach offered a new and somewhat more comprehensive definition. He defined evaluation broadly ". . . as the collection and use of information to make decisions about an educational program" (Cronbach, 1963, p. 672). This began a new movement in the field of educational evaluation.

Since that article, others have taken up and expanded upon this notion, producing most notably the CIPP Model of Evaluation, originated by Stufflebeam and Guba (Stufflebeam, 1967a, 1967b, 1969). This definition of evaluation is typified in the following:

> Project operations or activities are evaluated to influence decisions which influence program operations which are in turn evaluated, ad infinitum (Guba & Stufflebeam, 1968, p. 20).

Stufflebeam (1969) also writes:

> . . . evaluation means the provision of information through formal means, such as criteria, measurement,

and statistics, to provide rational bases for making judgments which are inherent in decision situations (p.53).

These viewpoints are representative of those in the literature dealing with the relatively new notion of educational evaluation as being decision-maker oriented. Taken together, they represent what can be called a Decision-Maker Model of Educational Evaluation.

Another basic notion needs to be brought up at this point: What is, or who are, decision-makers? A decision-maker is that person or group of persons who are responsible for making decisions regarding an educational enterprise. Or, from the perspective of the evaluator, the decision-maker(s) is/are the person(s) for whom data will be collected and to whom the collected data will be reported for the purpose of assisting or aiding the decision making efforts.

In the Decision-Maker Model, the actual, in-fact project personnel are the decision-makers and further, their role as decision-makers is legitimized in this Model. That is, this approach to educational evaluation assumes these things, among others:

1) That the project or enterprise decision-makers, be they classroom teachers, principal or superintendent (all of whom are potential decision-makers) have the right - both morally and ethically to make their own decisions about their own enterprise.
2) That it is the responsibility of the project or enterprise decision-makers to make their own decisions. It is not the responsibility or the right of an outside "expert" or "consultant" to do that.
3) That the only legitimate purpose of educational evaluation is to provide information to these decision-makers for their own use as they see fit.
4) That the validity of this approach is determined in the final instance by whether and to how great a

degree the data are used by the decision-makers in making their decisions.

There are a number of other assumptions that separate this approach from the more traditional ones. First, it assumes that decisions can be made more effectively with appropriate data. Implicit in this purpose is that data, to be appropriate, must come from the decision-makers' individual project, not from some external sources; and furthermore, that the decision-makers involved must believe in and be ready to use the data that are to be collected. Thus, evaluation takes on a new relevancy when based on internal needs, wants, criteria and data rather than on the outdatedness and ineffectiveness of the application of external (and therefore probably unrelated) standards and criteria to a project.

This conception also demands that the decision-makers involved have the final say in the determination of what data they want and need to make the kinds of decisions they deem important and necessary, not data defined solely by an evaluator, or data determined by arbitrary external criteria.

It is assumed further that evaluation is not a one-shot, post hoc procedure, where if the tests show you have succeeded by 90% you can sit back and relax, patting yourself on the back (although not knowing where you succeeded and where there is still room for improvement) or conversely, if the tests show you failed, e.g., achieving only 20%, you groan and chalk up a lost year, still not knowing where you failed or what parts if any are working. To be effective, evaluation must be built into a program from the first so that the constant and continuing decisions which need to be made during a program can be made on the basis of data wherever and whenever possible, rather than on impressions or intuition alone.

Finally, it demands that before any data be collected, the decision-makers involved need to know not only what data they want, but also what data they need and will use, why they want it and how they are going to use it. In other words, they must define the goals of their project in order that appropriate data may be gathered. Notice here also that this is an internal problem, not an external one.

An evaluator's job within this framework of evaluation is to assist the decision-makers in stating goals, in deciding what data are to be collected and how they might be collected. An evaluator's job is not to dictate which goals are important, which goals should be chosen, what is "good" or "bad" and so on.

This approach to evaluation is essential to decision-makers who are concerned with how well they are doing by their own standards, where they are failing and so on. This approach does not tell the decision-makers what decisions to make, but rather only shows them where they need to be made.

II. THE FIRST STEP IN EVALUATION

At this point, some decision-maker in the enterprise makes the decision (and follows through on it) to have an evaluation done. He contacts an evaluator and sets up an initial meeting. What kinds of things should be expected at that first meeting? What should the decision-maker look for? What should he ask and expect to be asked? This section of the paper focuses on these questions.

Negotiation of the Contract: Initiation of the Evaluation

The purpose of this first meeting between the evaluator and the decision-maker who has been responsible for having the meeting set up is to develop the scope of the work for the evaluation. What kind of decision-maker would organize such a meeting? It could be the assistant superintendent who has been asked by a group of teachers, or the superintendent or some other decision-maker(s) to contact an evaluator. It might be a team leader or a principal who feels a need to have an evaluation done and so proceeds to contact an evaluator. In short, it could be any decision-maker who has some legal and financial ability to bring in an outside person to do work, in this case evaluation work.

Assume now the evaluator has come to a meeting with the project or school or enterprise. What happens? The decision-maker should expect to be asked the same question posed a few pages earlier in this work: "Why do you want to evaluate?" The purpose of asking this is to make sure that it really is evaluation that is needed and wanted and not something else. If the purpose is to provide some kind of data for decision making, then the majority of educational evaluators practicing evaluation today will probably continue the discussion. If some other purpose is given, then the evaluator might (probably) try to help the decision-maker specifically define that purpose and then suggest another type of consultant who might better help achieve that purpose (e.g., a PR man or a needs analyst).

Following agreement on the purpose of evaluation, the next likely thing to happen is for the evaluator to begin to explain

what he or she can and can't do in terms of an evaluation. The decision maker at this point should look for what tasks will be accomplished, by whom, and so on. The decision-maker should feel free to ask any questions that might be bothering him and clear up any confusion he feels.

If at this point both the decision-maker and the evaluator feel comfortable with their respective positions, then the discussion will get more specific, or at least it should get more specific. The decision-maker should expect to be asked something like "What it is that you want evaluated?" The evaluator might also be concerned with what the purpose of the enterprise is: how complex it is, i.e., are there many parts and decision-makers involved or is the enterprise small enough to be viewed as a single project or program? If the evaluator feels that the enterprise is too broad or too vaguely defined, he will probably try to help the decision-maker narrow it down.

For example, an assistant superintendent has invited an evaluator to an initial meeting. He says:

"I want my school system evaluated."

The evaluator sees this description of the enterprise as somewhat broad and responds:

"You want the <u>whole thing</u> evaluated?"

The decision-maker responds:

"Well not the whole thing, but the reading program."

Again, to make sure this is the enterprise to be evaluated, the evaluator might ask:

"The whole reading program, system-wide?"

"Not really, just this new reading curriculum we have in the Model Elementary School."

In other words, the evaluator wants a fairly explicit description of the enterprise. He would probably go on to ask what are some of the major elements of the program; some of the major concerns, etc. He might ask for a brief description in writing. The decision-maker should expect such a discussion.

This initial meeting will also deal with <u>resources</u>. It takes resources to do an evaluation. Resources are defined as: staff time, secretarial and clerical support, duplication costs,

decision-maker time, and money. In other words, people usually think of resources as a fancy name for "money" but money is only part of resources. The decision-maker should expect then to have to identify the resources which will be made available to the evaluation. Again, this is going to probably be more than just quoting a dollar ($) figure. If the evaluator does not ask to have resources identified, then the decision-maker should raise such issues as:

Who is going to type up and distribute progress reports?
Who will pay for the phone calls back and forth?
Where will meetings take place between the evaluator and the staff involved?
Who will organize and convene these meetings?
Will there be a final report printed (if appropriate)? and who will do it? in how many copies?
Who will print data collection instruments?

These are just a few of the kinds of issues that need to be resolved during this initial meeting with the evaluator and if the evaluator does not raise these issues then the decision-maker had better or he is liable to find a lot of hidden costs appearing later on. Before the discussion concludes, then, the decision-maker and evaluator should agree on a list of resources, including all those things mentioned above in addition to money.

Another and perhaps more important issue which should be raised and resolved in this initial meeting (and which is often overlooked in many evaluations) is to identify for whom the evaluation is to be done. An evaluation can not be done outside of a particular context; in the absence of specific people. An evaluation is done for people who have particular needs for the information to be collected by the evaluation. (After all, the purpose of an evaluation to to provide information to someone for that person, or group of people, to use in making decisions.) In other words, who are the decision-makers of this enterprise who will be provided with data? At first glance, this question may seem simple and obvious: "Well, I called you Mr. Evaluator to come here so I am the decision-maker." Right? Not quite. The evaluator should respond with something like "Well, do you make decisions about the program we are going to evaluate?"

Well of course.

"Are you the only one?"

"No," the decision-maker responds, "there are the teachers in the program who make daily decisions."

"Is that all?"

"No, the principal also makes some decisions about it. For that matter so does the superintendent. If you start to think about it, there are a lot of people who make decisions about our reading program."

As it turns out, for any educational enterprise, be it as small as a single class or as large as all Title III projects in the country, there are many, many decision-makers and not just those usually thought of as decision-makers (e.g., administrators). For example, in an evaluation done of an experimental K-1, integrated day Title III project, decision-makers identified were: (1) the team teaching in the program (4 persons); (2) the principal; (3) the other teachers in the school; (4) the Superintendent; (5) the school committee; (6) the parents of the children enrolled in the program; (7) the Title III office in Boston. Each of these different decision-makers wants and needs different information to make their decisions since each makes different decisions from the others. To collect different sets of data or information for each decision-maker in the above example would cost a fortune! because each would require a different evaluation design. Thus it is not only important to identify decision-makers, but also to put them in some priority order since in all probability it will be impossible to pay to have an evaluation done for each of them and a single evaluation will not be appropriate for all of them at the same time.

Part of this discussion then should also provide for prioritizing decision-makers. There are any number of ways this can be accomplished but what is important is that it be made very clear to all parties at this initial evaluation meeting who will be getting information.

A related topic is how much of the resources which have been identified earlier in this discussion will be allocated to each decision-maker. That is, of the total amount of resources, how much will go to the evaluation for the first priority decision-

maker(s), to the second and so on. For the example given of the experimental K-1 program mentioned above, 100% of the resources were allocated to the highest priority decision-maker, the K-1 teaching team. It was decided, however, to report information collected for them to the other decision-makers but not to do an evaluation for the others. Resources just did not allow for such a wide ranged approach.

(It should be noted that providing data collected for the primary decision-maker to other decision-makers in the enterprise does not constitute an evaluation for those "others". Such data may or may not be relevant to these "others'" decision making needs and there would be no way of knowing if such data were to be really used by these others in their decision making. Thus, simply reporting data gathered for one specific decision-maker to other decision-makers within the enterprise is not "evaluation" for those other decision-makers.)

Remember, an evaluation can not be all things to all people. It has to be determined, at this meeting, what it will be (and do) and for whom.

This should just about cover what will (should) happen at an initial meeting between a decision-maker and the evaluator. Again, any doubts a decision-maker has should be expressed and dealt with; any misunderstandings should be cleared up at this meeting; both the decision-maker and the evaluator should feel comfortable with each other and with what each wants to do and can do.

Review: Of the initial meeting between an evaluator and decision-maker

(1) Have you discussed the purpose of the evaluation and come to a mutual understanding with each other?

(2) Have you specifically defined the 'enterprise' to be evaluated and come to a mutual understanding with each other?

(3) Have you had all your questions answered satisfactorily?

(4) Have you identified a list of resources which includes more than simply money, but staff time, secretarial support, materials, etc.?

(5) Have you identified the potential decision-makers of the enterprise identified in #2?

(6) Have you ordered these decision-makers as to whom evaluation data is to be provided?

(7) Have you decided what percentage of resources should be allocated to each decision-maker?

(8) Is the scope of work and responsibility of the evaluator and decision-maker (or makers if there are more than one) been clearly established?

(9) Has the time period for the evaluation been clearly established?

Each of these points or questions should be dealt with at an initial meeting between evaluator and decision-maker. This review section can be used by a decision maker, if he or she likes, to check or assess progress during the first meeting. In other words, this list can serve as a list of criteria for assessing what has or hasn't, does or doesn't occur during an initial meeting. A decision-maker can know what he has gotten and not gotten and act accordingly.

Preparation of the Evaluation Contract

The contract should be prepared following the initial meeting, just described above. It should include all the information used in answering the questions above, not just an agreement to do an evaluation. It should include the purpose, enterprise, resources, decision-makers, time lines, responsibilities; in short, all the topics agreed upon between the decision-maker and evaluator at that first meeting.

Once the contract has been prepared, it should be gone over carefully by both parties and both parties should agree and be comfortable with each point in the contract or the contract should be changed.

Concluding remarks to the decision-maker:

(1) Unless you are very satisfied with the contract and happy with its provisions, don't sign it until you are. Otherwise, you may have cause for regret.

(2) Don't accept the contract that simply says Mr. Evaluator and Model School agree to an evaluation for $X.XX. "Evaluation" is a fuzzy concept and can include (and exclude) many things. Before you sign a contract, be sure you know what you are getting and that you want, need and like what you will get. Simply, who will do what. There are responsibilities required from each party involved in the evaluation process.

III. A GOALS PROCESS

Whenever an evaluation is done, it should have as one of its steps some kind of goals process. The purpose of such a goals process is to identify those intents, or aspirations, or goals which the enterprise being evaluated is to accomplish. If the evaluation is to collect data, on what is it to collect data? The answer to this question is: on those goals the enterprise is to accomplish.

The goals process is a very important part of evaluation. It provides for the selection of variables as well as providing the basis for designing the entire evaluation. If the goals process is incorrectly applied, then data to be collected later will be less complete, less efficient and less focused than it should be. These three factors in turn will cause the evaluation to be less effective than it should be. In short, there can be no efficient evaluation without a systematic, reliable goals identification and priorization process.

Goals occur on all levels of specificity and do not have attached to them the rigorous criteria of specificity prescribed for behavioral objectives by Popham and Baker (1970), or Mager (1962). Table I lists some of the possible differences between the two classes of phenomena. Goals embody intents, the intents of the decision-maker, not just the verbalized, specific statement of what the decision-maker thinks his behavioral objectives are.

Because of recent trends in education, it is important to clarify terminology. For example, it is important to distinguish between the concepts of "goals" and "objectives". This is a crucial distinction to understand. The use of the word "goal" is intentional. The popular catchword in education today is "behavioral" or "instructional" objective. However, there is a distinct difference between the "goal" concept and the "objective" concept, which is, or should be, a subset of the goal concept.

TABLE I

SOME POSSIBLE DIFFERENCES BETWEEN GOALS AND BEHAVIORAL OBJECTIVES

GOALS	BEHAVIORAL OBJECTIVES
1. General, vague, not very specific.	1. Specific behavioral verb.
2. Fuzzy; may overlap with other goals; may be in conflict with other goals.	2. Single specific verb object, excluding possibility of overlap.
3. Embodies real intents.	3. Reflects writer's ability to write behavioral objectives.
4. Does not really communicate specifics to others.	4. Communicates very well and specifically to others.
5. May be stated in terms of anybody, including inanimate objects.	5. Stated in terms of the learner.
Examples:	
1. to have individualized instruction.	1. The student must be able to correctly solve at least seven simple linear equations within a period of thirty minutes.
2. self-actualization	2. Given a human skeleton, the student must be able to correctly identify by labeling at least 40 of the following bones: there will be no penalty for guessing (list of bones inserted here).
3. autonomous learner	3. The student must be able to spell correctly at least 80% of the words called out to him during an examination period.
4. open classroom	

(These are taken from Mager, 1962, pp. 45-50.)

Rather than asking the decision-maker to write down all his behavioral objectives, as many "traditional" approaches to evaluation would ask, following which the evaluator would then proceed to "measure" their achievement, a different tack is called for. This different tack is necessary for several reasons. First, the former approach assumes certain behaviors, skills and knowledges on the part of the decision-maker: (1) the ability to write behavioral objectives; (2) the ability to translate the decision-maker's purposes or intents into meaningful behavioral objectives; (3) the ability to write objectives embodying all his intents. To assume these skills on the part of any decision-maker is both illogical and potentially damaging to the overall evaluative effort. (For further discussion of this subject, refer to Hutchinson and Benedict, 1970; Benedict, 1970).

The decision-maker is asked what he would like his "enterprise" to accomplish, the word "enterprise" being defined as that entity about which data is to be collected. (An enterprise can be a school, project, class, program: that which is to be evaluated).

This approach, using an interactive relationship between decision-maker and evaluator should yield an initial list of "goals". The most noticeable quality of this initial list is that these "goals" are usually vague or nebulous. Differentiated staffing; educate good citizens; graduate responsible Americans: all of these might be typical of the level of specificity of goals at this initial level. Even though they are stated as fuzzy concepts, they embody real intents and aspirations on the part of the decision-maker.

It should be pointed out that fuzziness is not always "bad". It is "good" in the sense that it serves the purpose of allowing people to operate in the ordinary communication process of the day-to-day world. People communicate in fuzzy concepts; they dream in terms of fuzzy concepts and they aspire in terms of fuzzy concepts. If these fuzzy concepts are avoided by going immediately to behavioral objectives there is the great risk that the behavioral objectives that are identified will not add up to the full set of the decision-maker's aspirations.

What is important, then is that the elicitation of goals be as complete as possible, whatever they may look like grammatically. It is essential that the evaluation begin with all the goals. Otherwise there is the possibility of missing or omitting what might be some of the most important intents of the decision-maker for the project. (Beginning with goals is possible because a methodology does exist for dealing with the fuzzy concepts in goals: the Operationalization process discussed later in this paper).

A goals process should have at least three major provisions: (1) mechanism for generating a list of items or goal statements; (2) a mechanism for insuring the completeness of the list; and finally, (3) a mechanism for ordering (or prioritizing) the list of goals.

Generating a list of goals: The evaluator should elicit the decision-maker's goals, being very careful not to insert into the process his (i.e., the evaluator's) own goals, nor his own interpretation of the decision-maker's goals. Beware the evaluator who debates a decision-maker's goals with him; who tells the decision-maker what he (the evaluator) thinks the decision-maker's goals should be. If the evaluator "forces" a goal on the decision-maker which the latter really does not want or does not hold, then data collected on that goal will not, and cannot, be used for decision making and the evaluation will either be incomplete or fail entirely depending upon the extent to which this "forcing" occurs.

Insuring completeness of the goals list: As pointed out earlier, one of the purposes of a goals process is to arrive at as complete a list as possible of decision-maker intents. The test of completeness mechanism helps to achieve this purpose.

One of the criteria of evaluation is that the data provided be "complete", and the notion behind a test of completeness stems from this concept of "completeness" in evaluation itself. Completeness in evaluation means that (within the resources available) all the data a decision-maker needs to make his decisions is provided to him by the evaluation. To insure this, at each of many decision points throughout the evaluation it is necessary to

"test the completeness" of many different processes. By doing this throughout the evaluation, rather than at say a terminal point, the evaluation design becomes more complete; data provided to the decision-maker will also be more complete.

The thinking behind how a test of completeness works is basically this. A decision-maker, in being asked to think of a certain class or set of phenomena, may spend an hour or two doing just that. However, this causes him to have a certain psychological set about those phenomena, or, he becomes "locked into" a certain pattern of thinking. To ask him to keep thinking in this same pattern is not useful for he has probably exhausted the process from that perspective. A test of completeness is meant to jolt him out of that set or pattern by offering or stimulating the decision-maker with a different perspective, a different set of phenomena, to which he may react. After having him get into this new pattern by reacting to a set of phenomena from a different perspective, he would again have a certain psychological set. And, depending upon resources at the various points of the evaluation, he would then be presented with yet another set of phenomena from a different source and so on. It is very important, then, that the evaluation have some provision for insuring completeness of goals. Such tests of completeness should not be the evaluator's own goals but should come from within the decision-maker's enterprise.

Ordering the goals list: Once the goals list has been generated and tested for completeness, it is necessary to put it in some sort of order. This list may contain anywhere from one to one thousand goals. It is impossible physically (and financially) to proceed with an evaluation on twenty or thirty fronts at the same time. It is necessary to proceed at one point. A Prioritization mechanism provides for a systematic ordering of the decision-maker's goals such that the evaluator will know how to proceed. It is very important that the decision-maker decide this order (with the evaluator assisting him in an objective and systematic fashion). The evaluator should not determine this himself.

Review: A Goals Process

When an evaluation is being done, does it do or have the following:

(1) use the decision-maker's goals?

(2) ensure that the goals are really those held by the decision-maker?

(3) ensure that the evaluator does not interfere by inserting his own goals or feelings?

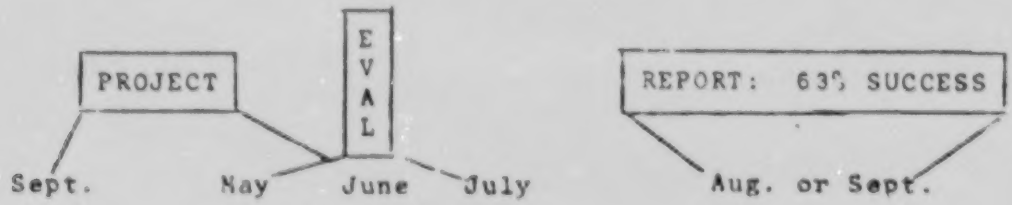
(4) that as many as possible decision-maker goals are identified?

(5) that there is an ordering process of some kind that results in an ordering that is acceptable to the decision-maker?

IV. A PARTS PROCESS

Unless resources - including time, staff and money - are extremely limited, an evaluation design should have as one of its steps a "parts" process. What does this mean?

One type of evaluation information or data one ofen sees looks something like this:



The evaluation is done near the "end" of the project.

We might term this a post hoc evaluation procedure where some sort of measurement or testing is done at the end of the project. This is a one-shot type of evaluation.

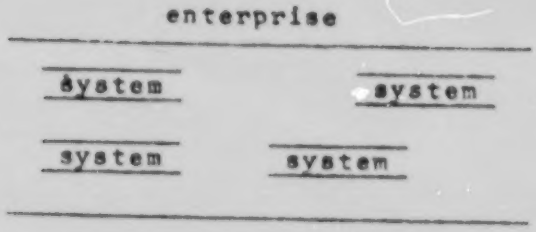
But, the next question is "So What?" What usefulness is there in deciding that the enterprise is doing well or poorly or is 63% satisfactory? What decisions can decision-makers make on the basis of this? If the report shows 80% success in June, does the project pat itself on the back and applaud? What if the report shows only 20% success? Does the project then wring its hand and chalk up a whole year to failure? Furthermore, what was 80% or 20% successful anyway?

In short, such information is of little utility in knowing what succeeded or failed. The utility of evaluation should be in knowing what parts or components or elements of the enterprise are working well and which are not working very well, and in addition, knowing this at the time it is happening when there is time to correct it, rather than after it is all over.

One needs to be able to assess each part or component as it contributes or fails to contribute to the purposes (goals) of the enterprise.

Instead of looking at the enterprise as a whole,

We look at the components or parts or systems of the enterprise.



If one has the parts of the enterprise, one can evaluate each part as it contributes to the goals of the enterprise. The purpose of a parts process is to identify the parts of the enterprise from the point of view of the decision-maker for whom data is to be collected.

One can find what isn't working which can provide the basis for making change and evaluate the change - thus one is freed to innovate because one can really know whether or not the innovation is better.

Again, this is in keeping with the idea of providing continual data to decision-makers for the purposes of making the continual decision any project must make. One needn't and shouldn't wait until it is all over and then either shout or cry.

How might this be done? The evaluator should work with the decision maker to identify the parts of the project being evaluated.

This is not as difficult as it may sound. Every system has a certain number of givens, i.e., given elements. Among these are Input, Interfaces and Output.

Input: those things occurring before the enterprise begins, or those pre-requisites for the program. Examples in a school situation might be budget, a physical plant and so on.

Interfaces: those things which are not directly a part of the project but which impinge on it and thus influence it. Examples again in a school situation might be the School Committee, parents, PTA, Legislature and so on.

Output: that which results from the project or program, that occurs after the program is ended. In a school, the output might be the student after the program or at the end of the year.

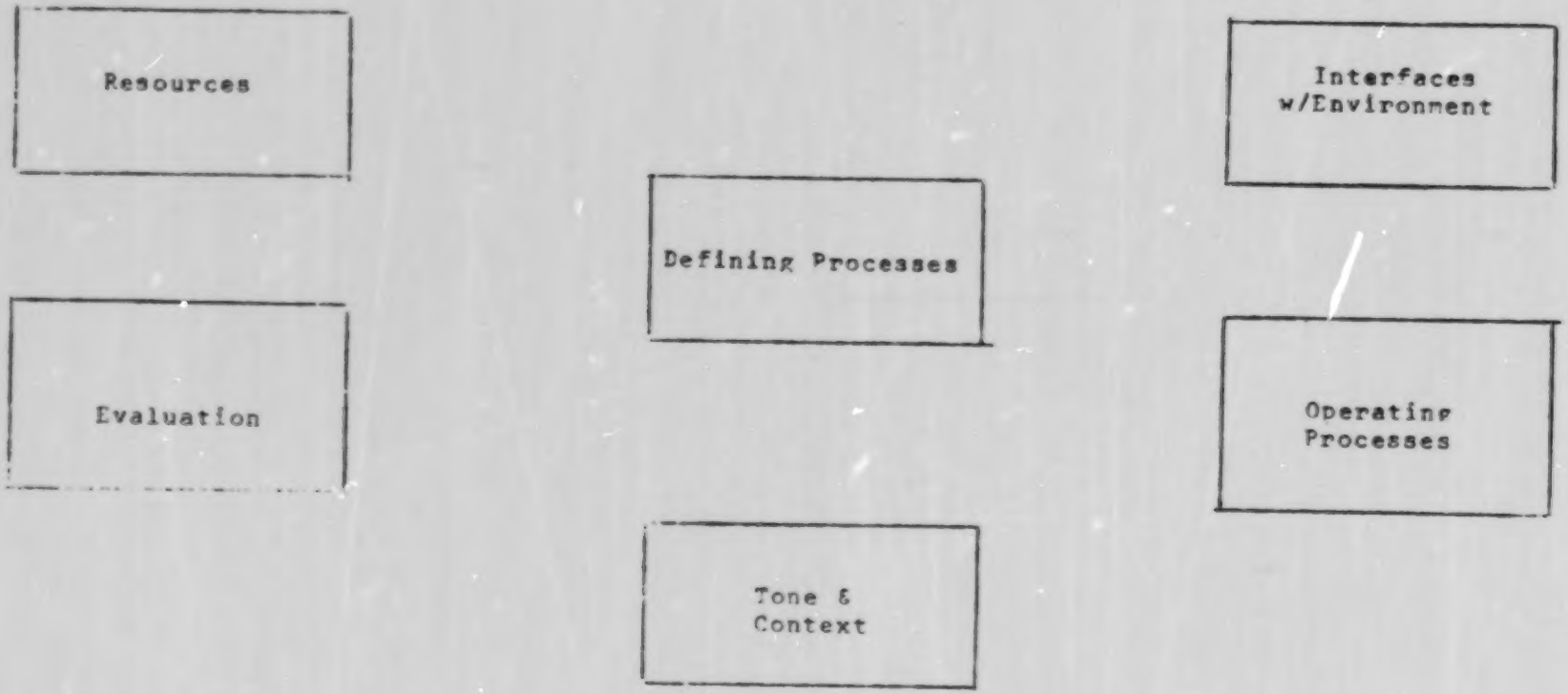
Now, for evaluation purposes, what is needed is the decision-maker's conceptualization of what these systems of the enterprise are. The decision-maker should be asked to list the major conceptual components or parts of the enterprise. For example, "When you think of your enterprise, what are the major things (parts) in which terms you think of it?"

The evaluator should not tell the decision-maker what his (the decision-maker's) parts or systems are. He may tell the decision-maker about Input, Interfaces and Output as general categories, but the evaluator again should not fill in the content of the categories for the decision-maker. The evaluator should also not give the decision-maker too many examples because the evaluation design might end up with someone else's, not the decision-maker's, components. If this were to happen, the evaluation will begin to lose its efficiency.

Several other points should be made here about a "parts" process. Different decision-makers may and do conceive of the same enterprise (or system) in different ways. (Example I) shows components of a school of education from the perspective of the Dean (a decision-maker in such an enterprise). (Example II) shows the components of the same school of education from the perspective of a School Council (another decision-maker in the same enterprise). These two examples show how a single enterprise can be viewed very differently by different decision-makers within it. A third example (Example III) is also provided, which shows the components of an Early Childhood Program from the perspective of the teaching team (the primary decision-maker in this particular enterprise).

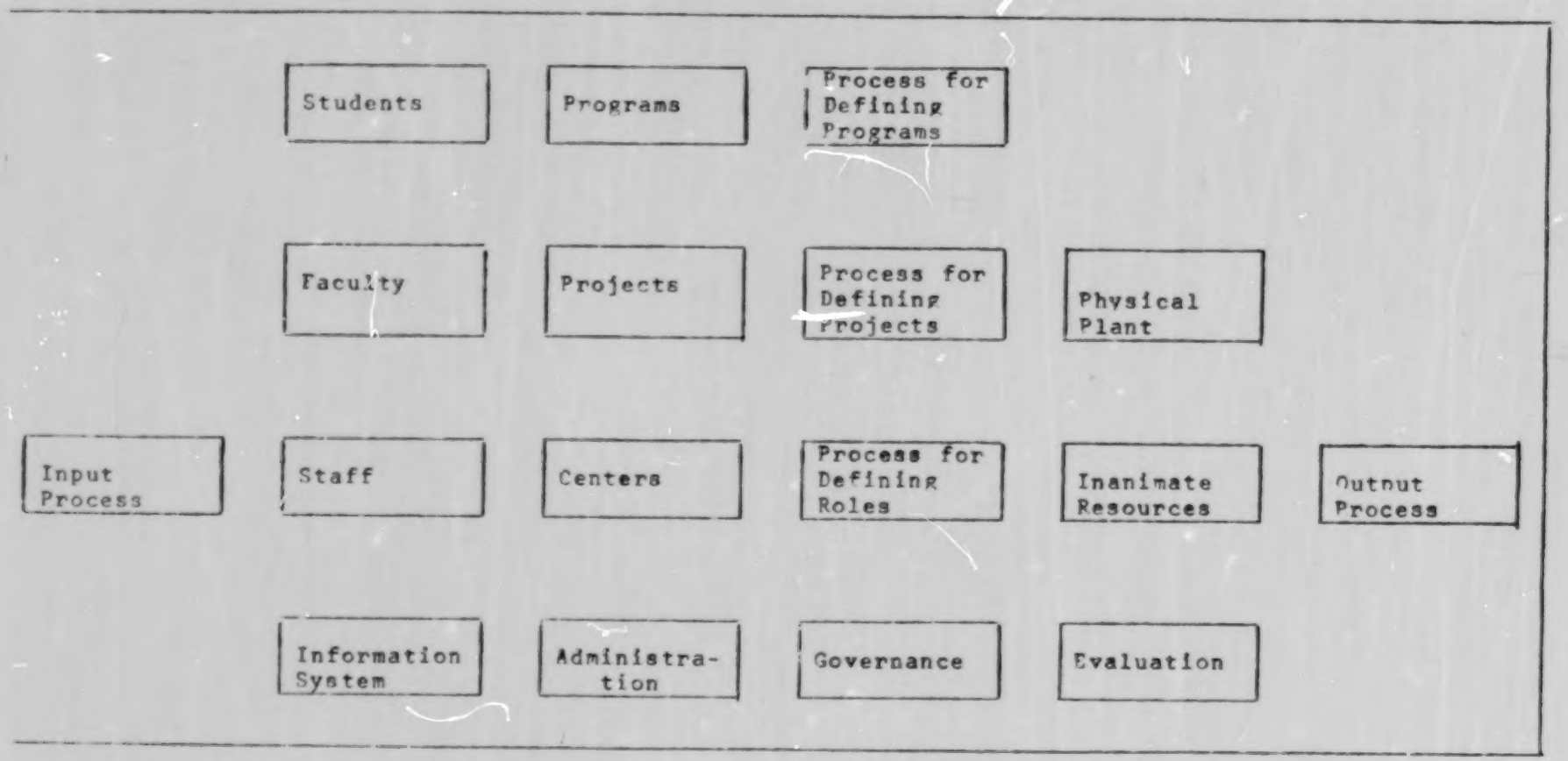
In the three examples given, the enterprise has been broken down one level. Conceive of the enterprise as a whole as level 0 of breakdown. Once the major parts of this have been identified, consider these the first level of breakdown. Each of the systems at the first level of breakdown are in themselves systems. As such they have input, output and interfaces, and other subsystems.

A SCHOOL OF EDUCATION



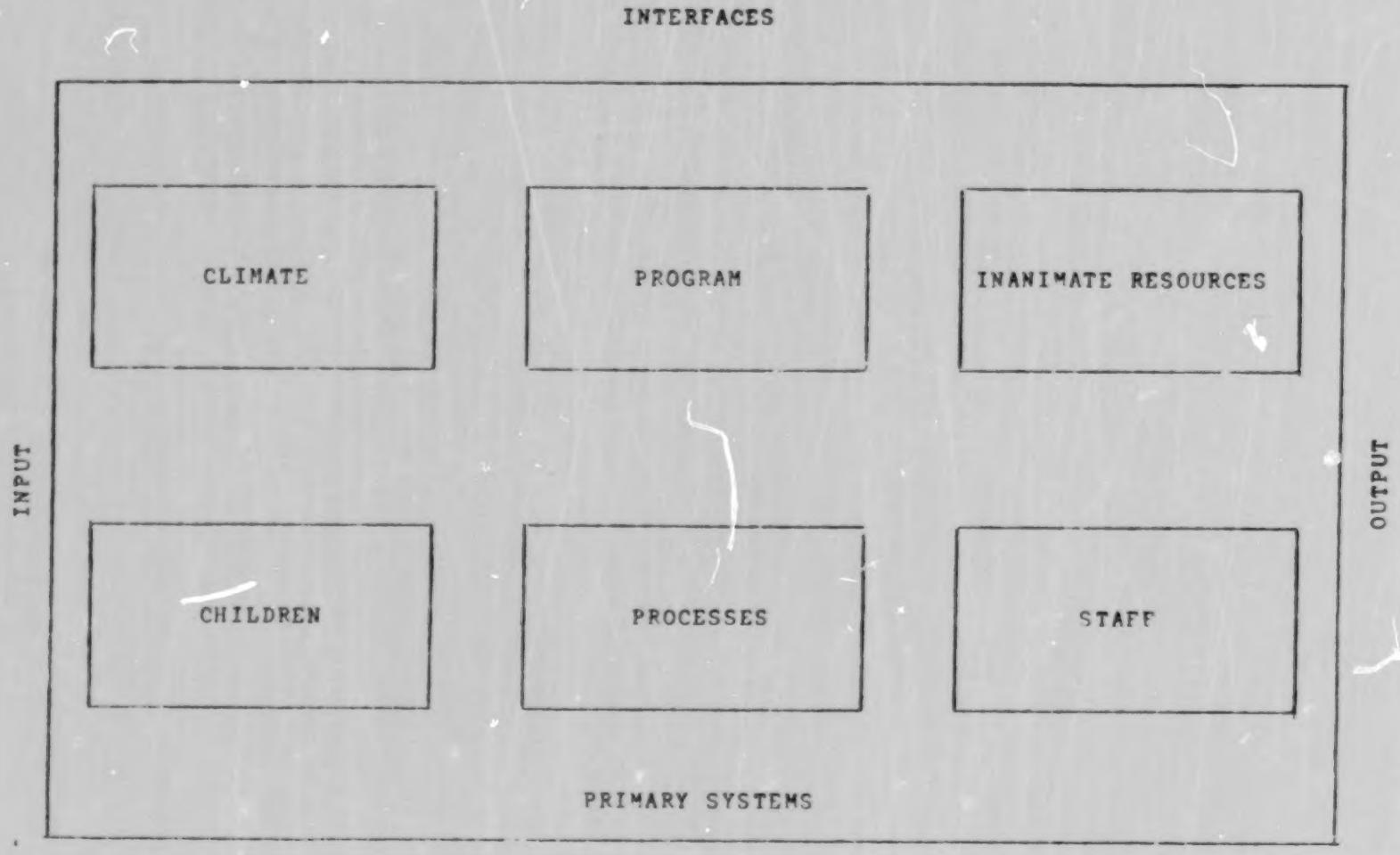
Example I. First Order System: Dean's

A SCHOOL OF EDUCATION



Interfaces with Environment

Example II: First Order Systems: School Council's



Example III: First Order Systems: Mark's Meadow K-1 Enterprise

The next step in a parts process is to go to the second level of breakdown for each of the systems identified at the first level of breakdown.

For example, look at the system labeled 'Climate" in example IIIA. Climate is the first level of breakdown from Example III. In this instance, when broken down one more level, i.e., the second level of breakdown, two sybsystems were identified: 'Physical Climate' and "Affective Climate."

An evaluation design should provide then for some kind of 'parts" process, from the perspective of the decision-maker for whom data are to be gathered. The parts process, like the goals process, should have at least three major provisions: 1) a mechanism for identifying (or generating) an initial list (or set) of parts; 2) a mechanism to insure that all the major parts have been identified, and finally, 3) a mechanism for matching goals to parts since the original purpose of parts was to be able to evaluate the enterprise in terms of its parts vis-a-vis goals, not the whole enterprise.

The purpose of the first mechanism and what it might look like are described in the beginning part of this section. In terms of the second mechanism, as with goals, the objective here is as complete a systems breakdown as possible. The more complete and specific the analysis of systems, the more specific and meaningful data can be related to specific parts of the project and not the project in its most global sense.

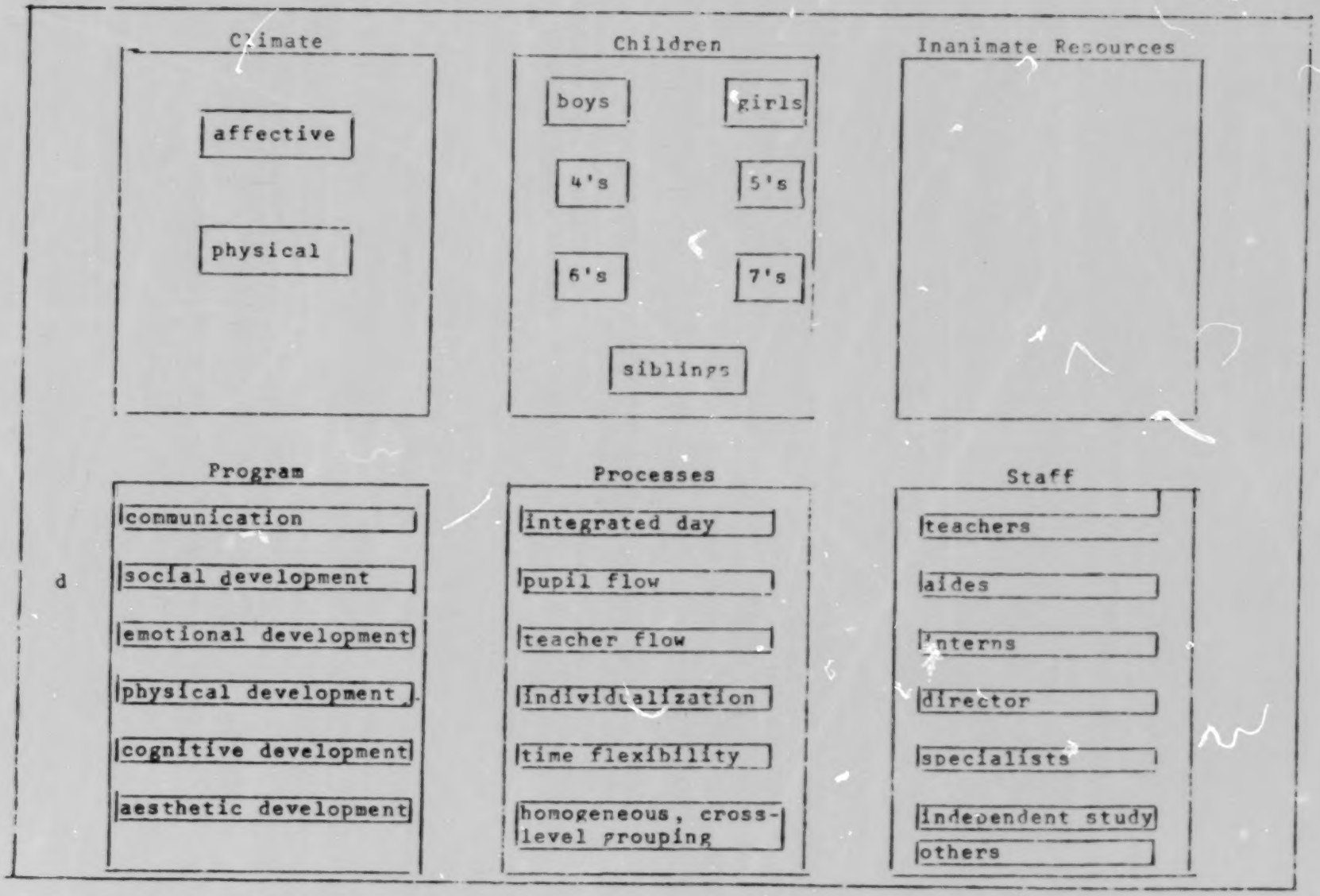
Concluding Remarks:

Do NOT be alarmed, or frustrated, or depressed and throw up your arms and say "I'll never be able to do all this". You're not supposed to - the evaluator is. This material is being presented here so that when you hire an evaluator, you will know the kinds of things to look for, to expect and the purpose of these processes. This material is also being presented here so you will have some criteria against which to measure, or gauge, or evaluate" the evaluator and the evaluation.

Is evaluation complex? Yes, it is.

Is it easy? No, it is not.

This material it is hoped, will better allow you to go into an evaluation with your eyes open, knowing what to look for, a little less anxious than you might have been. Evaluation is meant to help you and if it doesn't, then it, the evaluation, is not working and needs to be improved. You are the decision-maker: the evaluator is the evaluation expert.



Example IIIA: Second-Order Systems: Mark's Meadow K-1 Enterprise

Review: A Parts Process

Then let's review this section as to what to look for in an evaluation:

(1) Does it have or make provision for providing data in terms of parts of the enterprise?

(2) Do the parts come from the decision-maker for whom data is going to be collected? (They should.)

(3) Are there mechanisms for generating a list of parts? for insuring the completeness of the parts list (or diagram if you prefer)? for matching the goals to the parts? (There should be.)

V. A MATCHING PROCESS FOR GOALS AND PARTS

Once the goals have been identified from the goals process, and the parts have been identified from the parts process, there is a need for a process to relate goals and parts to each other. A prioritized list of goals has (or should have) resulted from the goals process and a prioritized list of parts should have resulted from the parts process. Now, these need to be matched to each other. This is done because of the purpose of doing a parts analysis in the first place; to increase the efficiency and usefulness of the data which is to be provided for decision making.

One way of doing this matching job is shown in the example diagrammed on the next page. The enterprise in this particular evaluation is a high school course in mathematics and the decision-maker in this particular instance is the teacher of that class. The goals, listed in the left column were his (the teacher's) goals for the enterprise and the parts on the top row were also his.

Wherever an X appears in a box, it indicates that the goal in the column is supposed to be accomplished, at least to a degree, by that part, or system, of the enterprise. Each and every goal should relate to at least one part. Each part should have at least one goal related to it. Such a diagram makes it possible to observe if there are goals for which no part has been identified to fulfill them. (Is there a goal and no "X's" in the row next to it?) This example does not provide an instance of this occurring but should it occur, it would indicate a need to the decision-maker relative to the design of the enterprise.

Such a diagram also makes it possible to see if there are parts without any (seemingly) useful function. (Are there any parts under which there appear no "X's"?) Again this example does not provide an instance of such useless parts but should such have appeared, it would have indicated a need to the decision-maker relative to the design of the enterprise.

The Evaluation should not tell the decision-maker to make a decision or that a decision is needed. The Evaluation would simply provide data and point out any discrepancies (such as the two possible cases described above of missing parts or useless parts) and leave any decision making up to the decision-maker.

ENTERPRISE: HIGH SCHOOL MATH COURSE

PRIMARY DECISION-MAKER: TEACHER

GOALS \ PARTS	Teacher	Students	Class work	Home work	Interaction with other students	textbook
1. know set theory	X	X	X	X	X	X
2. illustrate basic concepts of set theory with Venn diagrams	X	X	X			
3. ability to translate word problems into Venn diagrams	X	X	X	X		
4. ability to translate Venn diagrams into word statements	X	X	X	X		
5. to enjoy math taught in this fashion	X	X	X	X	X	X

Review: Goals/Parts Matching

(1) Does the evaluation have a provision for somehow matching the goals of the enterprise with the parts of the enterprise for the decision-maker? (It should.)

(2) Does this matching process use the goals identified from the goals process and the parts identified from the parts process for a given decision-maker? Or does it use one decision-maker's parts and another's goals? (The latter shouldn't happen.)

(3) Does the matching process provide for the decision-maker doing the matching? (It should.) Or, is the matching done by the evaluator? (It shouldn't be.)

VI. AN OPERATIONALIZATION PROCESS*

This is one of the most important processes within an evaluation. It deals directly with the problem of translating what a decision-maker wants to do, into an observable or measurable state. It is also an area where such current evaluation models as Stufflebeam's CIPP (Context, Input, Process, Product) Model, Provus' Discrepancy Model and the EPIC Model fall far short of an ideal and in fact, do not satisfactorily deal with it at all.

After all these years, there is still a dichotomous trend in education with vapors to behavioral objectives. On the one hand there is Mager (1962), Bloom (1956), Popham (1969), and Popham and Baker (1970), all of whom represent a school of thought which would have us detail in minute, behavioral terms the objectives of whatever it is we are about, or else, they pose, we'll never know where we are going or where we have been. On the other hand, there is an increasing movement with spokesmen like Atkin (1963), Ausabel (1967), Raths (1968) and Eisner (1969) which questions the efficacy of the former school, suggesting that when forced to operate along Magerian lines, the essence of what we are about may very well be lost, or that the behavioral objective approach is limited in its ability to deal with things that are really or should be of concern and importance to us, e.g., affective goals. Despite Popham's (1968) excellent refutation of this latter point of view, an uneasiness still remains with us about the efficacy and desirability of one or the other of these two seemingly polar opposite points of view.

These two positions may not be polar opposites. The problem may be that our abilities of conceptualizing are still in too immature a state to handle the non-Magerians versus the Magerians points of view simultaneously. The point is:

---

* The majority of this section originally appeared in, Hutchinson, T. E. and Benedict, L. G., The Operationalization of Fuzzy Concepts, University of Massachusetts, mimeo, September 1970.

> Evaluators, educators, all human beings, have enormous difficulties in reporting the sum and sweep of their objectives. We all have goals and we consciously and unconsciously give priority to some goals over others. But we have few reliable ways to report them to others, or even to reveal them to ourselves. (Stake and Denny, 1969, pp. 375-376)

This is the crux of the matter. We all have goals but getting from goals to verbalized or explicit statements of what these goals mean not only to others but to ourselves is the problem.

For example, it is easy to state, "The student shall solve 5 quadratic equations in 5 minutes without the use of any materials other than scrap paper and a pencil." It is easy to communicate this to others with full understanding, as it is an easy task to determine whether, if and when this objective is accomplished by the learner. However, this is not the case with a whole host of other kinds of goals, e.g., affective. "The student shall be self-actualizing. . .," or "The student shall value his self," and so on. These latter goals are difficult to communicate and understand and yet a legitimate argument can and is made that these are important as is solving 5 quadratic equations. Yet, while verbalizing these humanistic or affective goals, teachers and educators and objective-writers have failed to deal effectively with them precisely because their conceptualizing abilities have not been advanced enough nor comprehensive enough to do so.

Where is the solution? Can there be one? Is it true that without Magerian objectives we can not progress anywhere? Is it true, as the non-Magerians state, that putting content or goals into Magerian terms destroys that which is to be measured?

To date our conceptualization strategies have been limited. A possible bridge from the Mager to the Atkin position, i.e., a possible solution to this dilemma, may have been developed by Hutchinson (1969a, 1969b) - perhaps quite accidentally while working on solutions to other problems. He may have come up with a process whereby both the Magerians and their opposition will feel not only comfortable with what they are doing, but with each other. They need not seem to be polar opposites any longer, nor mutually exclusive, since in reality (it is contended) they are simply different points on a single continuum.

Examine for a moment some of the beginning of this controversy. Why is it that objectives ever began? It could have started when evaluation or assessment of student achievement began. It really came into focus with programmed learning with which Mager was really concerned when he wrote his book. The problem actually had its basis in the need for measurement. And this is the point at which evaluators entered the scene.

Evaluators and evaluations have had and continue to have a bad name. They are associated with anxiety on both the teachers' and students' parts. They have too often been part of the first school of thought mentioned earlier: "Tell me your specific behavioral objectives and then I will evaluate" is typically assigned as coming from an evaluator. As Stake and Denny write (1969), "An evaluator's technical skill should help the educator convey his purposes, both those that quickly come to mind and those implicit in what he does. What are the present methods . . . Our methods now are crude, unstandardized and unvalidated. They should be more evocative, more sensitive than indicated by the bold request, Please state your objectives in the following space." (p.376)

However, the above is not the only shortcoming of evaluators. A second is that of the subjective approach to evaluation. all too common a practice today. In this method of evaluation, the evaluator enters the situation and "feels" what is happening, or tries to sense some sort of global dimensions of what's happening, after which the evaluation is written. The problems with this approach are all too obvious.

Yet a third dimension which contributes to the fear and anxiety associated with evaluations is that the evaluator will use outside, unknown or irrelevant criteria to evaluate 'my school" , or "my course" or "ME". That this point has been compromised is evidenced, for example, by such criteria for a Social Studies Evaluation, as provided in the Natural Study of Secondary School Evaluation's, Evaluative Criteria (1960) as: enrollment, number of sections, range of class size, class periods per week, room arrangement and so on.

These problems with the current state of evaluations need not be the case. In fact, the whole nature of evaluation, what it is

and isn't, what it should and shouldn't do is changing (Stake, 1967, Stufflebeam, 1969, Scriven, 1967). Evaluation is headed for a new definition for which it indeed is time.

It is in this new movement of redefinition of the function of evaluation, and in developing a much-needed methodology of evaluation consistent with this movement that Hutchinson has devised a procedure he has entitled "The Operationalization of Fuzzy Concepts". An initial reaction to such a title is probably scepticism followed by "What is it?" Upon investigating this procedure, one discovered an extremely wide range of potential possibilities and applications. One such application is dealing with educational goals that are not easily turned into behavioral objectives.

What is a Fuzzy Concept?

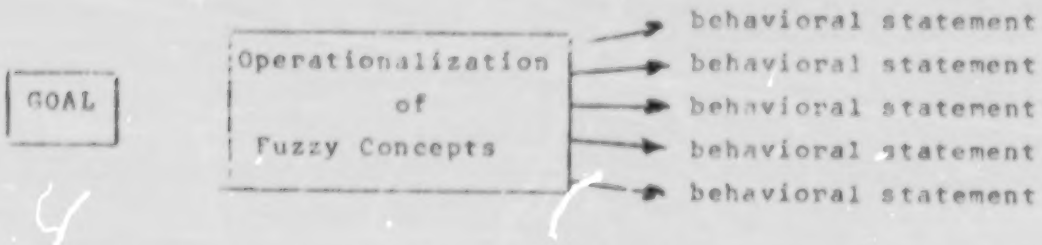
Fuzzy concepts are common. We all use them everyday of our lives in communicating: peace, love, democracy, patriotism and civil liberties are just a few examples of some of the many, many fuzzies used frequently today. Because each of us has different perceptions of the same words, such as those above, or phrases like self-actualization, individualizing instruction and student-centered learning, there often arises misunderstanding, disagreement, tension and even conflict. Often one hears the point made that what is really at issue is a semantic problem, a communication gap. This is due in part to the use of fuzzy concepts.

Fuzzy concepts can also be said to represent the dichotomy between instructional or behavioral objectives and goals, or non-instructional objectives. This very important difference or differentiation between goal and objective should not be underemphasized, overlooked nor confused. A goal, for example, is an "end" in non-behaviorally defined terms, such as "The student shall be self-actualizing". An instructional or behavioral objective on the other hand is an operationalized goal, e.g., "The student shall list in writing at least 5 directly observable components of his self-concept as he perceives it."

The apparent gap between the two schools of thought on the objectives controversy, between "goals" and "behavioral objectives",

is due in part to the fact that in reality these represent two different points on a single continuum, not two different continua. As Stake and Denny wrote, mentioned above, all of us have goals. It is simply a lack of conceptualizing strategies, an absence of a means by which to show that this gap is only an apparent one that is the issue in this controversy.

Hutchinson's technique, the operationalization of fuzzy concepts, may be the conceptual tool needed to resolve the issue. Keeping in mind the definition of goals, this might be represented as shown in Figure I.



A goal, when the operationalization technique is applied, will probably yield many behavioral statements (or objectives). It is important therefore not to dismiss goals, just as it is important not to dismiss objectives. The premise here is still the use of objectives, or operationalized goals. What is important is the way or means by which teachers and other educational decision-makers are exposed and introduced to the logic and necessity of objectives, as well as the way in which evaluators go about arriving at behavioral objectives.

Please note: the best way to learn this technique is to experience it. In order to maximize this experience the reader is asked to practice each step of the procedure as it is introduced and discussed. To simply read through this section trying to do each step will not be very effective for the reader.

The Operationalization of Fuzzy Concepts: A Methodology

Step 1: The first step in this procedure is for you to choose the fuzzy concept to be operationalized. Some examples are: peace, love, helping others, job satisfaction, self-fulfillment, etc. The reader should choose a fuzzy concept that he uses,

or intends to use, rather than one which is not important or meaningful to him. For purposes of this paper perhaps it would be easier if the concept "helping others" is used. Write the fuzzy concept on a piece of paper.

Step 2: Create in your mind a hypothetical situation. This hypothetical situation will have a group of people in it, an environment, things, furniture, etc. It may be indoors or outdoors. Now, imagine that the fuzzy concept exists in this situation and is in the epitome, is absolutely 100% present. Observe that situation and all the things you see about it that indicate to you that your fuzzy concept is present in this situation. The hypothetical situations should be as complete and real as possible. For example, the hypothetical situation in this case might be a classroom with chairs, tables, blackboard, etc. There is a teacher present, a group of students and so on. The teacher's behavior is the epitome of "helping others". List those things you can observe in this situation that indicate to you that the fuzzy concept is present. Some things might be:

a. concerned with the student as an individual
b. warm
c. sincere
d. considerate of students' opinions, values, etc.
e. smiles a lot
f. provides a supportive climate
g. provides success experiences for students
h. provides experiences for students to reduce their anxiety
i. provides experiences for students to define and reach their own goals

Obviously there are many others. Possibly none of these would appear on your list of your concept of

"helping others". Now, you should write your list down. Use this hypothetical situation completely, try to identify all the elements of "helping others".

Step 3: Now <u>again</u> <u>construct</u> a hypothetical situation and again with the environment and furniture, things, etc., a group of people and there is present in this situation the complete absence of the fuzzy concept, e.g., absolutely no "helping others" present. What things do you see in this situation that indicate to you that your fuzzy concept is completely absent from this situation. Let's take again the same hypothetical situation as was set up in Step 2: a classroom, a teacher, a group of students, etc. This time, imagine that this teacher is directly opposite the ideal of helping others. List those things you can see in this situation which definitely indicate to you this teacher is <u>not</u> "helping others". Some examples might include:

a. ignores students' opinions and values
b. not aware of students as individuals
c. egocentric
d. selfish
e. does not allow for individualization
f. authoritarian
g. discourteous
h. undermines students' feelings, morale, etc.

Obviously, again, these are only a few possibilities. Again, maybe none of these will appear on your list or fit <u>your</u> conception of "helping others". <u>Write down</u> all those things in this situation that you observe that indicate to you the fuzzy concept is absent. Don't bother with the negative statement of the positive elements listed in the previous step. Concentrate on identifying those aspects that were not already found.

Step 4: After having gone through both the positive and negative hypothetical situations, the chance of easily finding more dimensions out of one's mind is not very great. So next we employed some strategies called tests of completeness. (first test of completeness): Get someone else to go through the same steps as above with the same fuzzy concept. One then looks at the other person's list and considers item by item if the item should be on one's own list and if it is, add it to the list. Should you decide the item is inappropriate, reject it, i.e., it does not fit your conception. Or a third possibility is that the other individual's item may make you think of one or more dimensions you have forgotten (recommended perhaps because you dislike their dimension.) Ideally this test of completeness should be done with three or four other people.
Write down the appropriate dimensions which result from above.

Step 5: (second test of completeness): Go back and re-create the hypothetical situations. Now, there were things that you saw in those hypothetical situations that you wrote down, i.e., your two lists. There were other things that you saw that you did not write down. Go back, look again at those things that you saw and did not write down, and seriously consider the implications of these not being dimensions.
To use an example out of the context of "helping others", consider fuzzy concept "job satisfaction". If a person were operationalizing "success in a job", one of the dimensions which he rejected in the first hypothetical situation might be money. Now the question should be asked, "What are the implications for success in a job where the job provides

no money at all?" Suddenly it becomes obvious that for almost everyone money must play some role however slight in job satisfaction. So the dimension money is added, but perhaps a qualified amount, e.g., $10,000.

Now consider those dimensions you rejected for your fuzzy concept and write them down on your list if on reconsideration they are for you, a part of the concept.

Step 6: (third and last test of completeness): The task here is to deliberately construct some dimensions that have nothing to do with your fuzzy concept, in this case 'helping others', and again, consider the implications of these dimensions for your concept. Try that and in fact, write them down. Start out by asking yourself, "What has nothing to do with _____ (fuzzy concept)" and then, "Does it really matter?"

The example of our teacher "helping others" provided us with a number of dimensions of this concept. Now, did you consider the teacher's family life? relationship with his or her peers, the administration? Probably not, but is it not possible that each of these could have serious implications on that teacher's "helping others". The purpose here is not in fact to find things that have nothing to do with your concept but rather to attack the problem from a different perspective.

As you proceed through these steps, each one will be more difficult as the dimensions that comprise <u>your</u> conceptualization of what you mean by your fuzzy concept become more and more complete the number not identified become fewer and fewer and therefore hard to find.

After one has gone through the 6 steps in sequence, it is reasonable to conclude that one has a fairly

complete list of the parts of the concept at the just level of breakdown. This product of this process, then, might be represented in Figure 2.

Now using our example of helping others, as a result of the first 4 steps, some 17 dimensions of "helping others" were arrived at. Thus on the first level of Figure 2 there are 17 numbers. The next step in the process is:

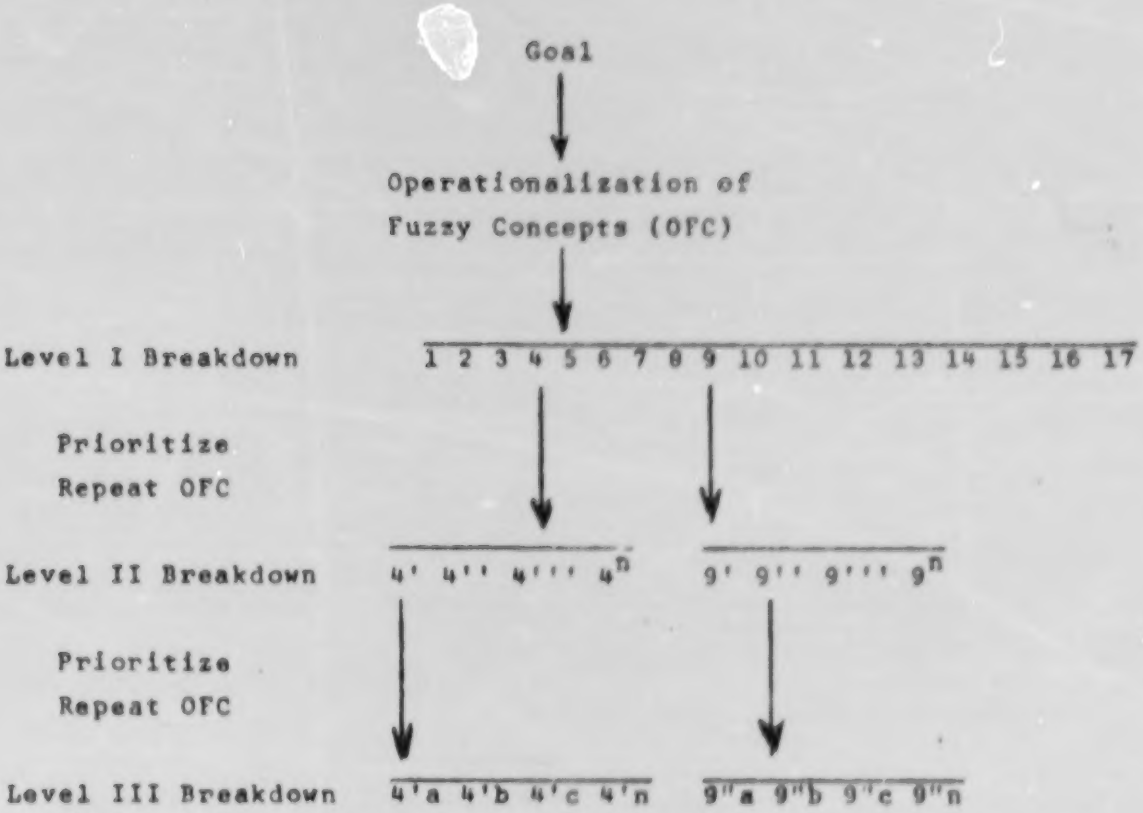
Step 7: For each item on your list, in this case 17 perhaps added to as a result of the tests of completeness, the reader should ask himself, "Can I observe that dimension directly?" Something which can't be observed directly is defined as a fuzzy concept. Thus, for each item you must decide if it is still fuzzy and if it is, then you must repeat, in the same order, the sequence of steps above.

In this particular example, none of the 17 items are directly observable and thus each must be further operationalized at least another level. Obviously at this point it becomes clear that this can be a very lengthy process. It could take nearly forever to do a complete operationalization. Thus at this point in the process, another technique is used, namely prioritization.

Since time is a resource and all resources exist in limited amounts, the reader must decide how much time he can allot to operationalization, depending on the reason he began the process. As an example, let's assume time is limited to a given amount and the operationalizer decides only items 1, 2, 12, and 14 can be operationalized. He repeats the process for each of these, including the important Step 7. Again, if an unmanageable number of dimensions are found each of which needs further operationalization, the prioritization at level two may take place, as in level one.

For a very fuzzy concept, what usually happens is that very few items at the first level of breakdown will be directly observable. As the operationalization process is carried further, a larger percentage are found to be directly observable.

FIGURE TWO


Goal
Operationalization of
Fuzzy Concepts (OFC)
Level I Breakdown
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
Prioritize
Repeat OFC
Level II Breakdown
4' 4'' 4''' 4^n
9' 9'' 9''' 9^n
Prioritize
Repeat OFC
Level III Breakdown
4'a 4'b 4'c 4'n
9''a 9''b 9''c 9''n

Perhaps it would be appropriate here to use a less fuzzy concept, one which can be fully operationalized in several levels rather than a large number. A fuzzy concept for a college physical education teacher might be "competent weight lifter". At the first level of breakdown, there are two dimensions: olympic lifts and power lifts. Asking the question, are these measureable or observable directly, the answer is "no" and the process is continued.

At the second level of breakdown, 6 more components are found, three from each of the first two: press, snatch, clean and jerk; and bench press, squat and dead lift. Further operationalizing "competent", certain attributes are attached to these dimensions, thus the third level of breakdown:

For a weight lifter with a body weight of 123½ pounds or less

press: 150 lbs.
snatch: 150 lbs.
clean and jerk: 200 lbs.
bench press: 200 lbs.
squat: 250 lbs.
dead lift: 450 lbs.

Each of these can be observed or measured by numerous methods and thus no longer fuzzy. The lifts themselves are operationalized by the current A.A.U. Weightlifting Handbook. (See diagram on next page.)

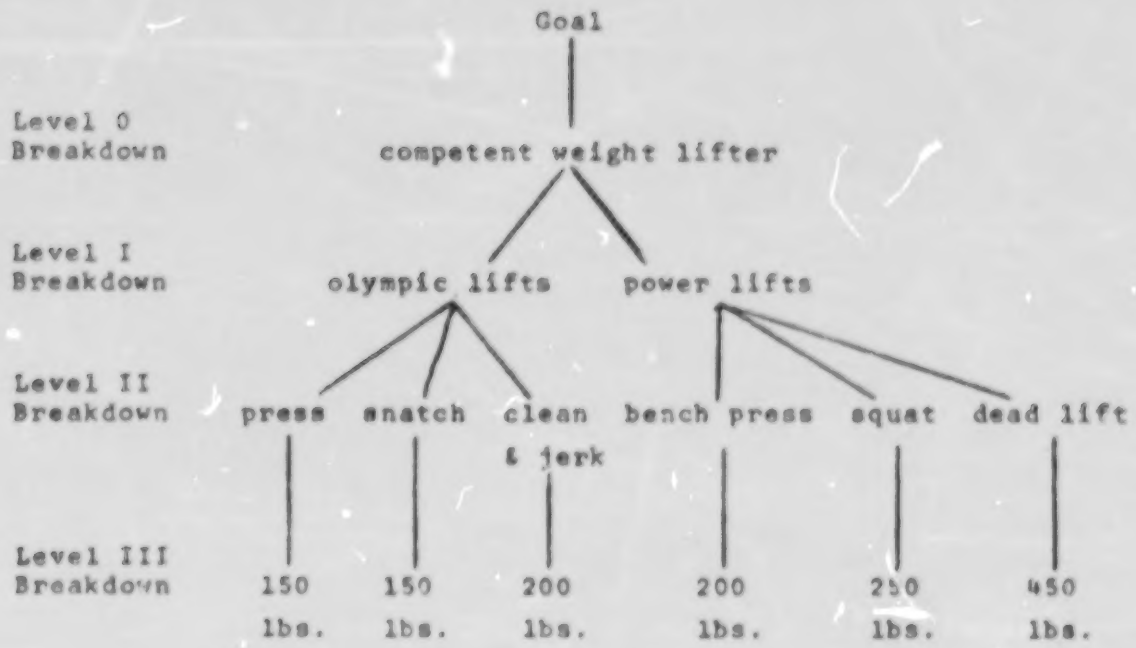
This was obviously a simplistic fuzzy concept with appeal to a limited audience. However, it exhibits how the process can and does work.

This then has been a brief overview of the operationalization of fuzzy concepts. It was introduced by two potential applications: first, as part of a new methodology of evaluation and second, as a method of resolving the objectives controversy.

An operationalization process should do the following:

1. Deal with the most important goals of the decision-maker for whom the evaluation is to provide data.
2. Take the most important goal and systematically break it down into behavioral, measurable dimensions or components.
3. Once the most important goal has been broken down, it will deal with the second most important goal and so on.

FIGURE THREE



4. Once operationalized, a goal or intent will consist of a whole list of observable or measurable items as in the weightlifting example.
5. These observable items should be prioritized by the decision-maker (with the evaluator's help if necessary).
6. Each item now becomes the behavioral item for which measurement for evaluation will be done. In other words, each item becomes the focus of developing a measurement technique which is then implemented and data collected.

The results of operationalization, then, form the basis for developing measurement techniques. This is the reason for the importance of the process. If the operationalization does not work, then data collection will fall far short of an ideal or best and may even fail completely.

Review: An Operationalization Process

An evaluation should have some kind of operationalization process. It may not look exactly like the one described herein. It may look entirely different. But, there has to be some sort of operationalization process. This is essential because of the need to break goals or intents into measurable, observable, behavioral statements. Merely starting with "write behavioral objectives" omits much that is important in terms of what the decision-maker wants to accomplish. Therefore an evaluation which starts with 'behavioral objectives' is falling far short of the 'ideal' and the decision-maker should be aware of this.

VII. MEASUREMENT FOR EVALUATION

Obviously one of the most important parts of evaluation is the collection of data. Data are collected using various observational techniques. The decision-maker for whom data are to be gathered and reported has a very important interest in the techniques which will be used to collect data. Therefore he should be involved in the development and/or selection of such techniques.

If the purpose of evaluation is to provide data for decision making; and if the data provided are to be used by the decision-maker; then any techniques used to collect data must be perceived as valid by the decision-maker or he will not use the data.

For example, if an evaluator is hired and he proposes to use a standardized test his concern or company has designed, the decision-maker should carefully examine it to see if it looks to him, the decision-maker, as though the information it will collect will be useful, that he will be able to use it. If the decision-maker feels that most of the information the instrument will collect will be useless to him - "It measures things I am not doing" - then it should not be used. Rather, a tailor-made instrument or technique should be used.

Most educators have had, at one time or another, a course in basic testing or in tests and measurements. Two concepts that most educators remember are "Validity" and "Reliability". Probably no two measurement concepts have been as referred to, or over referred to, in evaluation as these two.

What is validity? A technique is valid if it accurately measures what it intends to measure. For example, using a ruler to measure the width of a room is a valid technique. A ruler measures what it is supposed to measure: distance.

There are many kinds of validity but one of the most important, and the one most frequently overlooked in "evaluation" is "decision-maker validity". Decision-maker validity simply means: do you, the decision-maker, think that the data collection device suggested by the evaluator will collect the data that you want collected? that will be of use to you? In other words, do you,

the decision-maker, perceive the instrument as being valid (measuring what you think it is supposed to measure)? If the answer to these questions is "Yes", then the technique or instrument is said to have decision-maker validity. If you, the decision-maker, are skeptical about an instrument or measurement technique; or have doubts about its ability to do what you want it to do, measure what you want it to measure, then the instrument or technique is said to lack decision-maker validity and should not be used.

What is reliability?

Does the technique perform consistently with time? For example, if we had a ruler which expanded several inches on a hot day or contracted several inches on a cold day, it would not be a reliable measurement technique because it would not perform consistently each and every time it was used. A technique has to be reliable (consistent) or it should not be used.

An instrument can be completely reliable and very 'valid" in the traditional testing sense and yet supply completely irrelevant data to the decision-maker for whom it was intended to collect data. In the past, traditional tests, testers and evaluators have concentrated on 'validity' (not decision-maker validity) and reliability to the exclusion of the decision-maker's needs. (This is only one reason why so many "traditional" evaluations have failed, i.e., have sat on the shelf and collected dust.)

In terms of evaluation, when it comes to the measurement, the decision-maker should expect some interaction with the evaluator on the development and/or selection of a technique. If the decision-maker leaves this entirely in the hands of the evaluator, chances are very good to excellent that the data collected will not be completely useful for making decisions and possibly will be entirely useless to the decision-maker. There is the example of the outside evaluator hired to come in and evaluate a summer workshop whose purpose was to take pre-school, disadvantaged children and give them readiness activities in preparation for their entry into first grade. The evaluator arrived with two tests in hand, administered them, wrote up a report showing a few

significant differences, (mostly no significant differences) and sent the report to the decision-makers. The decision-makers reacted: 'Neither test measured what we were doing!' "We were dealing with emotions and attitudes and he (Mr. Evaluator) tested cognitive development'.

In this example, both tests had been field tested, were valid in testing terms and reliable but did not have decision-maker validity. As a result, the decision-makers rejected the whole evaluation, fired the evaluator and decided to find an evaluator who could develop and provide measurement techniques which could collect data about what they (the decision-makers) were actually doing.

In the first place, then, an observational technique must fit that which it is to measure. It must be developed or selected from existing techniques for a specific task: collecting data on a specific goal or intent which the decision-maker may hold for his enterprise. Prepackaged tests or standardized often fail to do this since they are usually on such a general level (in order to measure a wide range of things) that they miss collecting data on the specific needs of a specific decision-maker.

Part of decision-maker validity is determining, by the decision-maker, for himself, whether a technique seems to fit that which it is to measure. If an instrument is clearly going to measure cognitive development and the major concern of the decision-maker is psychomotor activity or affective components of that cognitive development, then regardless of how valid or reliable is that measure of cognitive development, it will fail in this instance because it does not measure what it is supposed to. It would not have decision-maker validity.

"But", the decision-maker is going to say, 'How do I (we) know about validity?" Sometimes it is just a feeling, an intuitive distrust based on experience, as with the example just given. However, there are a number of criteria a decision-maker can use to determine whether an observational technique is useful, valid, and going to serve his needs.

Criteria to assess observational techniques

The decision-maker can ask himself: is the technique direct observation of behavior or is it indirect observation. Direct observation is always preferred to indirect because it gives a much better indication of what is really happening. For example, if the item to be measured is "children fighting in the schools" it would be best to collect information by direct observation - counting the number of fights per day - than to give a self-report questionnaire to all the behavior problems in school asking them to write down the number of fights in which they have been involved. Students and non-students alike know how to "distort" answers on a written test to the direction the question asker wants. They know they are not supposed to fight so they report "no fights" when in fact there may have been several. In such situations direct observation is always preferable to indirect.

Is the technique obtrusive or unobtrusive? An obtrusive measurement technique is something which is not ordinary but which is introduced only for the "evaluation" so to speak. Obtrusive techniques share the same problem that indirect measurement had above: it interferes with that which is being measured and may very possibly alter it. For example, if the item to be measured is "cheating" (the peeking kind) an obtrusive technique is to have two or three persons stand in the room to watch for peeking. An unobtrusive measure might be to have a one-way mirror and to stand behind it and count the number of peeks. Unobtrusive measures are preferred where possible to obtrusive ones. Perhaps the best example is the annual or semi-annual trip by an administrator to "evaluate" the teachers. The administrator comes into a teacher's room with his checklist or pad of paper, sits glaringly or even smilingly in the back of the room busily writing. The teacher's behavior will automatically change for the duration of this "obtrusive" measure. Whether the change is for the better or worse is not the point: the point is, what is being observed is not what is usually happening because the obtrusive technique is interfering and interacting with that which is being measured.

A third criteria which can be used in assessing measurement techniques is that of naturalness. Is the observational technique to be used under natural conditions or under unnatural conditions (e.g., test)? That administrator was observing his teacher under natural conditions - her natural classroom environment but he violated one of the other criteria. Thus it is important to note that having just one of the criteria may not be sufficient. In the case of the teacher, perhaps again, observing through a one-way mirror would have been natural. (Granted, very few schools have such devices: remember, this is only for illustrative purposes.)

There are other examples of "unnatural" conditions which the decision-maker can be on the look out for in reacting to or assessing observational techniques: simulations, models, lab situations, test-taking conditions. Each of these is unnatural to an extent and will therefore distort to an extent that which is being measured.

An ideal observational technique then will be reliable and valid (especially decision-maker validity) and it will also fulfill three other criteria: direct, unobtrusive and natural. But, as with all ideals, it is very seldom met. Meeting all of these criteria will be both expensive (usually) and sometimes impossible. The ideal observational technique for determining certain behaviors of teachers, say, is an invisible man. This is obviously impossible although highly desirable in many circumstances.

However, knowing what is ideal, the decision-maker can then know how far from the ideal a given observational technique is. He can use these "criteria" of idealness to measure observational techniques the evaluator presents or develops. It becomes very useful, therefore, for a decision-maker to have a rough idea in his mind of what an ideal technique might look like for any given item to be measured.

These criteria become very important in the realm of the affective domain, psychomotor domain and in the areas of attitudes and emotions. In the cognitive domain, there has to be a strong reliance on paper and pencil tests (again, remembering though that even this is far from the ideal) but such "tests" are far from satisfactory in the other areas listed.

Review: Measurement for evaluation

(1) Have you, the decision-maker, been involved in the development or selection of observational techniques?

(2) Do the observational techniques have your "decision-making validity"? (That is do you feel the data collected by them can be used by you? Meet your needs?)

(3) Have they been field tested and been shown to be reliable?

(4) How direct is each technique?

(5) How unobtrusive is each technique?

(6) How natural is each technique?

(7) In short, how far from the ideal is each technique and is this so far that it loses decision-maker validity?

Again, these can be used as criteria by the decision-maker to know what he is getting or is not getting in the way of measurement in evaluation.

Beware: the evaluator who has one or two or even more pre-packaged tests which he plans to administer which you, the decision-maker have little or no say about. Such tests will probably not provide you with useful or useable information and therefore should be regarded with skepticism unless it can be shown that these are the very best available. (This can be partially answered by going through each of the above 7 questions with the evaluator and posing them to him.)

VIII. DATA COLLECTION

Once an observational technique has been agreed upon by both the decision-maker, who has certified that the technique has decision-maker validity, and by the evaluator, who has certified that he can use it and that it is usable in terms of testing validity and applicability, then that technique is implemented and data are begun to be gathered.

There are several criteria which the decision-maker should be aware of to use in assessing the process of implementing the technique. Granted the evaluator (or a measurement consultant who might be called in) has expertise in implementing observational techniques but there are certain things a decision-maker can also look at which allow him to make some observations or decisions about the implementation of these techniques.

First, when does the evaluator plan to collect data using a given technique? If the evaluator has planned to use a technique only once, at or close to the end of the project then the decision-maker should question the advisability of this. Data should be provided on more than a terminal or after-the-fact basis. The decision-maker should use some reference to his needs for data before accepting a suggestion to use a technique once, when the project is nearly over or the school year is nearly over.

How often should a technique be used? There is no exact or correct answer to this question. For example, the following is a goal which is held by a teaching team for their enterprise, in this case, a primary classroom:

In the room, many children's things are displayed.

The observational technique developed for collecting information on this is simply: to randomly pick a time during the week; send an observer into the classroom to count all things displayed which are children's things (not teacher things). (Children's things include: art, papers, things brought from home to show to the class, etc.)

It was decided to implement this technique for the first time in October of the year.

Time I: In the classroom there were 12 children's things displayed (drawings, sculpture, papers, etc.)

The primary decision-maker (the team of 4 teachers) decided that this was really not sufficient to meet their intents for this goal and so they decided they would work at increasing the accomplishment of this intent. In this case, the technique was used again a week later and this time,

Time II: 35 things displayed

The team decided that they had reached a satisfactory level on this and would now turn to other things.

This does not mean that the technique was never used again. It would be used again to see if this level were dropping off, staying the same or increasing (each of which would indicate a different set of conditions necessitating a different kind of decision).

Time III: (4 weeks later): 39 things displayed (all of which were incidentally, different from the 35 things seen 4 weeks earlier).

This confirmed the decision-makers' perceptions and feelings that this goal was being more than satisfactorily met. In this case the technique might not be used again for 2 months.

But, what if at Time III there had been only 10 or 15 things displayed, all of which had been on display when observed 4 weeks earlier? This would probably have caused alarm and would have allowed the decision-makers to deal with this in any number of ways, with any number of decisions. (Evaluation does not tell the decision-makers what decisions to make or what caused the conditions necessitating the decisions. Evaluation provides data to the decision-makers which they then use to make decisions or not, as the case may be.)

They immediately took action to correct the situation, made several changes in their program, etc. In this instance, the technique would be used again very soon, perhaps 1 or 2 weeks later.

In other words, this has all tried to say that how often a technique is used depends on the needs and decisions of the decision-makers. A decision-maker should then be wary of the evaluator who wants to simply give a post-test. Suppose in the

above example, a post-test were given in January or in June and it found that only 10 things were displayed. If school were out for the summer in June, it would have been much too late to do anything and it might have indicated that this particular goal had been inadequately met, in fact it had not been met at all. If it had been done in January, half the year had gone by, with a situation existing which really needed change. It is important, therefore, not to rely on such rules of thumb as post-tests. Seldom if ever will such data collected be of great decision making utility.

(Note again that in the example, given, direct, natural and unobtrusive measurement was done. A questionnaire was not given to the teachers to ask them what they did. Observation was carried out to determine it.)

Implementation of measurement techniques should reflect decision-maker needs and decisions made.

It should also be remembered that the frequency of use of a technique will vary from technique to technique, as well as for the same technique. Therefore the decision-maker should not expect all the techniques to be administered or implemented on the same time schedule or with the same frequency. This would not be efficient, or focused. Such a rigid pattern of collecting data would not yield the most effective information. (The most effective information is that which is there when you need it, in the amount you need it, and where you need it. Collecting all the range of information all the time as would happen if all techniques were used the same would not meet this definition of effective. In fact, such an approach to measurement is costly and a waste, both in time and energy and money.)

Exactly when and how often a technique is to be used is a flexible situation. The decision-maker who wants the most effective evaluation should expect a flexible schedule of collecting data and should raise questions if the evaluator wants to administer or implement techniques with the same frequency and in the same time pattern.

Sampling: Another criterion about which the decision-maker should expect to interact with the evaluator is that of sampling.

Sampling becomes a very important criterion when one reaches the stage of collecting data (implementing observational techniques). The evaluator should present any sampling plan or procedures to the decision-maker in order to determine whether the plan has decision-maker validity. The decision-maker should expect such an event to happen.

What is sampling? Sampling is picking a number of subjects from a larger group of them. For example, if there are 1,000 students in a school and one wished to determine how many were boys and how many were girls (assuming we didn't have this information) a sample might be taken all from the population (i.e., all 1,000 of them). This sample might be 10%. (it is cheaper to only deal with 100 than 1000 in terms of time, money, etc.) On the basis of randomly choosing a sample of 100, we find 55 girls and 45 boys. We might then, on the basis of this, estimate what the percentage of each sex is in the whole population, 55% to 45%.

This is a simplistic example to show that from a smaller sample, it is possible to estimate something about the larger population. If a population of students, or subjects to be observed is large, then some kind of sampling should be done in order to reduce cost. Observing all the subjects in a population is often expensive. This expense might be wasteful because sampling (when done scientifically and carefully) can yeld the same information, or a good approximation of it, which a census of the whole population would yield. In the 1972 national elections, a Gallup poll of only 1500 people was sufficient enough and representative enough to show what the whole voting population would do. In the sample approximately 60 or 61% said they would vote for Mr. Nixon. In reality, this percentage was almost exactly correct.

Sampling is done to save time and money and effort. Sampling is also done when it is impossible to find out a piece of information from all the subjects in a population (as in the example of the election.) There are two criteria within sampling which the decision-maker should look for: size and representativeness.

If one were measuring a goal on fighting in a school of 600 one would probably want to look at more than 6 students. A sample size of 6 from a population of 600 will probably be quite

inadequate. The size of the sample should be large enough that the decision-maker is willing to generalize from the sample to the population. Would a decision-maker generalize about 600 students from a sample size of 6? It is unlikely.

On the other hand, is it necessary to observe all 600 students to get an estimate of the amount of fighting going on in the school? Again, it is unlikely. A sample of students or a sample of classrooms will probably yield data which is valid enough to generalize to the school.

The sample size, therefore, should be large enough (or small enough) to maintain decision-maker validity without overspending resources. If the decision-maker feels that the data which will be gathered from the sample will reflect the actual level of goal attainment in the population as a whole, then the sample size is sufficient.

(There are certain scientific principles governing sampling and it may be that just decision-maker validity may not be "scientific" enough to justify certain generalizations. The decision-maker should expect the evaluator to point out such principles, in simple English during a discussion on sampling). However, if having to apply too many principles jeopardizes decision-maker validity to the extent that the decision-maker feels data to be gathered will be useless to him, then decision-maker validity has been "invalidated" and the decision-maker and evaluator need to discuss the problem. There is no sense in gathering data which no one will use in decision making.)

The second criterion the decision-maker should consider is that of the representativeness of the sample. Going back to the example of fighting in the school, it may be that the size of the sample has decision-maker validity, but that the representativeness of where that sample is to be taken does not. Let's say that the size has been determined to be 60 students. If the evaluator has designed a sampling plan whereby all these 60 students are freshmen, when the school has four grades, then this plan is clearly not representative. If, however, the goal was held for only freshmen, then a sample of 60 freshmen would be very representative.

If the sampling plan calls for selecting students from only social studies or only from industrial arts, when the goal is held for English also, then the plan is not representative. The decision-maker, then, should carefully judge whether the sample is going to be representative. If he feels it is not, he should raise this point with the evaluator.

In the final instance, it is the decision-maker who will use data for his decision making. It is the decision-maker who will have to generalize from data gathered from a sample to the whole polulation. To do this, he will have to carefully assess the size and representativeness of the sample.

Review: Collecting Data

(1) Is each technique dealt with individually with respect to how often and when it will collect data?

(2) Does the schedule for collecting data provide for flexibility such that this schedule can be changed (anywhere from more often to less often depending upon the nature of the data collected?)

(3) Has the evaluator discussed the sample and sampling procedures with you to determine your decision-maker validity?

(4) Are you satisfied that the sample to be selected is representative of the larger population?

(5) Are you satisfied that the sample to be selected is large enough to generalize to the larger population?

IX. HAVING EVALUATION DATA REPORTED TO THE DECISION-MAKER

"When is the data reported?" This very important question is one which is usually not addressed directly in evaluation and yet it is a crucial problem to consider. In many evaluations which have been done, the data are collected at one point in time and then the evaluator has cogitated, analyzed, summarized, synthesized, and interpreted the data all at the same time, following which he has written a report which is then delivered to the decision-maker quite often well after the need for evaluation data has passed, e.g., in August, three months after the project has ended at least for the summer. Or in September, two months after the in-service workshop has been conducted.

This problem of reporting data well after it is needed is one of the reasons evaluation has gotten a bad name and one reason that many people have criticized evaluation as being less than useful. What has to be done is that data need to be collected and reported as they are needed, not in one lump sum at some terminal point in a project or enterprise. In the previous section which discussed data collection, the point was made that in some cases the same set of data may need to be collected several times, especially when changes have been made in order to more likely reach a goal. Not to have the data reported until the end of that class year will mean that further decisions to make changes if they are needed can not be made and the purpose of evaluation immediately becomes less than being met. If data are not reported until the end of the year, for example, a decision to make a change or not to make a change can not be made on the basis of data. It is quite likely, in the example given of displaying children's things, that even the need for making a decision would not come into the open.

To be truly effective, then, data for decision-making need to be reported as closely as possible to when they have been collected. Also, the evaluator should be ready to collect the same data again in a short period of time if necessary. Data collection has to be responsive to decision-maker needs.

"What is to be reported? Again, this might seem to be a question with a very obvious answer but when it is considered carefully, it will be seen that it is really much more complex than is usually thought.

"The data are reported". This is the answer. But, what comprises the data? Data can be considered as the information gathered by the observational technique and they will probably have some numbers or figures or charts. This is what many evaluations report as data. It is really a narrow definition because there are many other things which should be reported in conjunction with these number "data' which become important in the decision making process.

A data report should include many things besides the numbers. It should contain the following things:

1. The name of the decision-maker for whom these particular data were collected. It has been pointed out many times that there are many decision-makers in an enterprise. If the primary decision-maker for whom these data were collected is the chairperson of the math department, then this information should appear on the data report. "Isn't this obvious?" one might ask? If is is, fine; if it is not, then it should be. The other decision-makers in the enterprise, e.g., the math teachers or the assistant superintendent for curriculum and instruction or the principals will probably, at one point or another also be given a copy of the data and it is essential that these other decision-makers know for whom and from whose perspective the data were collected. (Different decision-makers need different kinds of data. Reporting the data of the chairperson, to the principal if he does not know whose data it is, is likely to not view the data as meeting his needs. The point is, they may very likely not meet his needs because they were collected for someone else. This is why such labeling is important).

2. The name of the goal and its importance (or priority) to the particular decision-maker. Take, for example,

the earlier discussion of the goal 'having children's things displayed'. This intent was one of the operational components of the more general goal 'to have an affective climate in the program'. (The 'display' intent was only one of the many, many items. The data report for this particular item then should include the fact that this was part of the larger goal and that this larger goal was the #1 goal this particular decision-maker (the staff of 4) held for the program.

3. The importance of the operational component. The reader might be thinking at this point, 'But having children's things displayed', does not seem to me to be a very important part of 'affective climate'. The data report should also contain then the importance of the operational component to the decision-maker for whom it is being collected. For example, in this case, the report might contain "this component of display was ranked as number 27 of the 70 components of the goal 'affective climate'". This information then gives other decision-makers information for their decision making needs.

4. The name (and description if appropriate) of the observational technique used to collect the data.

5. The date of the data collection (or dates if appropriate) and the place, e.g., September 17, 22 and 28 in Mr. Teacher's class and Miss Teacher's class.

6. The actual data, presented in terms which the decision-maker for whom it is being collected can use and understand.

These 6 items are important items which should be part of a report on data. They are items which the decision-maker should expect. Such information clarifies the report and makes the data (in many cases) more effective, both to the primary decision-maker and other decision-makers of the enterprise.

Review: Data Reporting

(1) Is the data reported when it is needed? In the amount needed? On the appropriate items needed?

(2) Does the report include more than just a few numbers and statistics?

(3) Specifically, does the report include:
   a. the name of the person(s) for whom this particular set of data were collected?
   b. the name of the goal and the importance of the goal which this data is being collected to measure?
   c. the importance of this particular operational component to the larger goal?
   d. the name and description of the observational technique?
   e. the date, time and place of data collection?
   f. the data?

(4) Are the data presented in an understandable fashion? Such that they can be used and understood by the decision-maker for whom they were collected?

These are criteria a decision-maker should look for and expect in a report of evaluation data.

What a report of data should not have

Just as there are things which a decision-maker should expect and look for in a report on data, there are also things he should not find in such a report and if he does find such things, he should be skeptical about them and question the evaluator about including things which shouldn't be included.

The decision-maker should not find, within such a report, decisions made by the evaluator on the data. Decisions about the data, interpretation about the data, how significant are the data: these are properly made by the decision-maker. The evaluator should not write such things as 'These are good, the project should continue doing . . .' Or, 'These are bad, the project should change what it is doing and do this . . . Such conclusions and recommendations are outside the proper realm of the evaluator. Such inferences are for the decision-maker to draw.

The report should not contain evaluator biases in the form of passing his personal judgments about the data or the techniques or the observations. Such personal likes and dislikes of an evaluator are outside the scope of evaluation. (If a decision-maker wishes to hire someone who will come in and make such statements, then he should do so. However, such activity should not be called evaluation but judgment.

The report should not contain information from the evaluator which tries to influence the program in one direction or another; which tries to have specific or particular decisions made about the program's adequacies or inadequacies. These are in the domain of the decision-maker's responsibility. Again, if a decision-maker wants to hire someone to come in and make decisions, or recommend decisions then he should hire someone to do so, but he should not call it evaluation.

The report should also not contain a section entitled "Commendations" for the same reasons cited above. Many evaluation reports contain a list of things which are "commended" for the only apparent reason that the evaluator liked them. Such activities are outside the legitimate scope of evaluation.

The same can be said of a section in many evaluation reports entitled "recommendations". Such sections should be deleted for these are the responsibilities of the decision-maker. Everyone likes to be commended but many (if not most) decision-makers would argue with such "recommendations" which of necessity must reflect a shortcoming at least as seen by someone. A kindergarten teacher will not argue with those things she is commended for, but in at least one evaluation where the evaluator overstepped his bounds and included a section of Recommendations, the teacher, who was the primary decision-maker for this particular evaluation, disputed each and every recommendation with such responses as, "He doesn't understand kindergarten children", "He isn't an expert in early childhood", "He doesn't understand open classroom", "He recommends such and such which is not at all a goal of the program".

When an evaluator moves into the realm of "recommendations" and "commendations", he moves out of the proper realm of evaluation and into the realm of decision-maker for an enterprise of which he is not in fact a legitimate decision-maker. A decision-maker should beware the evaluator who want to, or does, get into this area of decision making for it is precisely that, decision making. Decision making is not evaluation. Evaluation should serve decision making and it can do this far better by not trying to coopt decision making but by providing data to proper and legitimate decision-makers.

Review: What a data report should not have

(1) Does the report have decisions (personal) of the evaluator? (It shouldn't.)

(2) Does the report have the personal likes and dislikes of the evaluator? (It shouldn't.)

(3) Does the report contain recommendations of the evaluator about the program, its direction, content, and so on? (It shouldn't.)

(4) Does the report have a "Commendations" section and a "Recommendations" section written by the evaluator? (It shouldn't.)

X. REDESIGNING THE EVALUATION

Redesigning the evaluation is an option which occurs only in certain circumstances. Ordinarily, the decision-maker would not expect redesign to be part of every evaluation but the topic will be discussed here so that the decision-maker might know what a redesign should include and when it might be done.

If the evaluation has been done properly to this point, with the interaction of decision-maker and evaluator and if the evaluator has been carefully fulfilling his role and not confusing his role with that of a decision-maker and if the decision-maker is fulfilling his role conscientiously, then there will probably be no need for a redesign section per se. Each step of the process, if the reader will remember, has a kind of redesign part to it. A step is not complete unless it has been satisfactorily agreed to by the decision-maker and evaluator. For example, during the goals process, the decision-maker must decide on which goals to include and which to omit. He must also decide on a priority order (with the evaluator providing the evaluation expertise necessary to help the decision-maker). If these processes are gone through and the decision-maker says, 'No, that is not the goals list I really hold,' or 'No, that is not the priority order of my goals,' then that particular section is recycled on the spot. This could be called a redesign of the goals process.

The same thing is provided for in each process of the evaluation. At least, it should be. A section is recycled or redesigned as a section until it is satisfactory. (Again, this is not likely to be necessary if the decision-maker has been actively and conscientiously involved in the evaluation design as he should have been).

What are some circumstances under which redesign of the entire evaluation might be needed?

Redesign might occur if or when:

1. The program or project changes dramatically or drastically. For example, the decision-maker within the project may leave, resign, die or be promoted, in effect changing the person(s) with whom the evaluator has been working

and for whom the evaluation has been designed. This would necessitate redesigning the evaluation.

2. The emphasis of the program changes (i.e., the goals change). During the course of a project or enterprise, goals are very likely to change. If this occurs, then redesign is necessary in order to reflect a change in goals or in priority of goals. This will in turn necessitate different observational techniques being designed, different data being collected, etc.

3. The enterprise experiences a 'break' or 'gap' between one part of its operation and another. This might occur in a Title III project, for example, which has been funded for three years. At the end of the first year, a decision might be made, or decisions made, which in turn would necessitate changing the evaluation. These decisions could deal with personnel change, program changes, financial changes, content changes, etc.

4. The enterprise is a long-term one. An example of this might be any part of a school system, e.g., math curriculum, English department and so on. In this instance, it is a sound idea to have an evaluation redesign stage built in. So many variables can change during the course of an enterprise, especially a long-term one that it really is necessary to provide for redesigning evaluation.

5. A conflict, misunderstanding or some similar problem, occurs between the evaluator and decision maker. This might happen for example if the two parties did not understand their purposes and functions during the first step of initiating evaluation and that misunderstanding did not become apparent until some time during the evaluation. Such misunderstandings could include or focus on: the purpose of evaluation, with one party wanting someone to make decisions and the evaluator designing an evaluation

to provide data to the enterprise decision-makers. Another example might be that in the initial phase of the evaluation, the wrong or incorrect decision-maker was identified. The decision-maker who actually makes the decisions was somehow not properly identified. This in turn would mean that the evaluation has been designed to provide data to the wrong person and thus a redesign would be necessitated.

6. Interpersonal relations-personality problems: As with any endeavor, these kinds of problems can enter the picture and could cause changes to be made. For example, the evaluator might have a value conflict with the decision-maker causing the evaluator to desire to leave the project. On the other hand, the decision-maker may experience value conflicts or personality problems with the evaluator and might cause him to ask the evaluator to leave. (A reminder might be made here that in preparing the contract, there should be stipulations allowing for this to occur without penalties to either party. A termination clause should be writeen in for the mutual benefit of both parties should the example just given arise. The decision-maker does not want to be saddled with a person who it turns out is completely incompatible with the needs of the decision-maker. Conversely an evaluator can not provide the most efficient evaluation design if he feels that there are incompatible differences between himself and the decision-maker.)

Review: Redesigning the evaluation

(1) Redesign may or may not be part of every evaluation.

(2) If redesign is necessary, it may be so for any number of reasons. It would be impossible to detail them all here. They are the same kinds of reasons which can cause problems in any educational enterprise.

(3) If redesign is necessary, then it should follow the same guidelines provided herein for a good evaluation.

(4) Finally, redesign is going to cost additional resources: especially time. The decision-maker should consider this before making the decision to have a redesign carried out.

(5) In the final say, it is the decision-maker who decides to have the evaluation redesigned or not.

Observation of the evaluation process by the decision-maker using these guidelines (provided throughout this booklet) may provide the basis on which to make the decision that a redesign is necessary. This could happen as soon as difficulty occurs in the evaluation process, rather than finding out during the last month of the evaluation that a redesign is needed. However, such a decision to redesign when difficulty arises can only happen if the decision-maker has been checking the process all along the way. It is suggested that the guidelines provided herein could serve as criteria to check the evaluation process throughout, not when it is done.

XI. EVALUATION OF EVALUATION

Evaluating the evaluation is part of evaluation. Yet very few evaluations which have been done have had provisions for evaluating themselves. In fact, most evaluations which have been done in the past usually terminate with a Final Report, when it is too late to systematically evaluate that Final Report.

One very important thing which a decision-maker should expect is to have some provisions made for an evaluation of the evaluation. As with all the other processes of evaluation which this booklet has discussed, the decision-maker must actively participate in this process.

If an evaluation is accomplishing its purpose, that is, providing valid data to the decision-maker for his decision making needs, then certain events are occurring and certain events are not occurring:

1. Data provided to the decision-maker are actually used by him (her, them) in making decisions.
2. The evaluation is efficient: All the data collected for a particular decision-maker are used by him. To the extent that data are collected and provided and not used, the evaluation has not met its purpose.
3. The evaluation is complete: Of the decisions made by a decision-maker relative to a particular program or enterprise, as many as possible are made with data provided by the evaluation.
4. The evaluation is focused: If data can not be provided (because of lack of sufficient resources like time and money) for all the decisions, then it should be provided for the most important decisions.

These three criteria - efficiency, completeness and focus - can be applied by the decision-maker to the evaluation in order for him to determine the extent to which the evaluation is meeting its purpose of providing data for decision making.

It is probably impossible that any evaluation will completely meet these criteria. There are many reasons for this. First,

evaluation efforts may be begun too late in the course of the program or project in order for data collected to meet the criteria. An evaluation can not fully meet the criteria if it is not begun until half-way through the project.

Second, resources will probably never be sufficient to allow the evaluation to completely meet the criteria. It is probably impossible to collect all the data, needed by all the decision-makers of a project to meet all their decision making needs, because the cost of doing this would be prohibitive. This implies certain things then which the decision-maker should take into consideration in evaluating the evaluation. The decision-maker must be cognizant of the amount of resources committed to the evaluation because resources determine the scope of the evaluation. He must remember that not all the data can be provided to all the decision-makers for whom it might be desirable. That is why, during the course of the evaluation, the primary decision-makers are identified and prioritized so that those persons most needing information might get it. That is why the most important goals of the primary decision-makers are identified so that they might get information on their most important needs or goals. If during the course of the evaluation even one of these was done incorrectly, the evaluation will become less efficient, less complete and less focused.

One way a decision-maker might collect information for himself so that he might evaluate the evaluation in terms of his own needs is to keep a log of decisions made relative to the program evaluated. Ideally evaluation and planning of the program occur at the same time, prior to the beginning of the program. If they are not or can not be, the decision-maker should remember that this will affect the evaluation of evaluation. For those decisions, he should note their relative importance to him. Then, he should assess whether and how much data was provided to him for those important decisions, and was it provided when he needed it. In other words, apply the three criteria.

What are some other things a decision-maker might consider in performing an evaluation of the evaluation? Evaluation should not interfere with the enterprise's accomplishing its goals (unless the goals are in conflict with one another and then this

becomes not a problem or fault of the evaluator but a decision making problem.) In fact, evaluation should help an enterprise to accomplish its goals by having information systematically provided during the course of that enterprise, such that the decision-makers of that enterprise can use it in their decision making.

Review: Evaluation of Evaluation

(1) Is the evaluation providing data for your decision making needs relative to the identified enterprise?

(2) Given the scope and resources of the evaluation:
- Is the evaluation efficient?
- Is the evaluation complete?
- Is the evaluation focused?

(3) Are you keeping a log of decisions you make relative to the identified resources in order to be able to assess points mentioned above?

(4) Does the evaluation or evaluator interfere with you and your enterprise achieving its goals? (They shouldn't)

(5) Finally, a person using this guide can evaluate the evaluation in terms of its parts, e.g., the contract phase, goals process, parts process, and so on, if he monitors the evaluation using the criteria provided for each section. This would be done in addition to keeping a log of decisions (in 3 above).

XII. WHEN RESOURCES FOR THE EVALUATION ARE REALLY SMALL, WHAT DO YOU DO?

This booklet has tried to present a comprehensive picture of the complex task of evaluation. However, the reader may have gotten the impression that "Well, this is all fine and good, but I have very few resources and I just can't buy all of this.

Resources will always limit the scope of the evaluation. Limited resources will have to limit the scope but do not have to exclude doing evaluation entirely. Limited resources simply will mean that the evaluation will have to be more efficient and more focused than unlimited resources.

The evaluation must in fact fit, from beginning to end, starting to deliver usable data within the resources that are actually available to do the job. Therefore resource allocation becomes a very important part of the evaluation. All the resources can't be spent on any one part of the evaluation, e.g., identifying goals, or doing a parts analysis. If resources are small, really small, then what is needed is as complete a goals process as possible within limits, as complete a parts process as possible within limits, and so on.

Limited resources will mean probably dealing with only one (the most important or primary) decision-maker of an enterprise. It will mean not doing a lot of tests of completeness in the goals process. Possibly, because of the focused nature of the evaluation (on a very specific and well defined enterprise) the parts process will be eliminated entirely.

Limited resources will also mean not operationalizing all the goals as completely as possible. It will probably mean operationalizing just the most important goal of the most important decision-maker. Throughout the evaluation, there will be short cuts and shortened forms of the processes. However, the basic processes should still be in the evaluation, even if in shortened form.

Even very limited resources will not mean that a decision-maker has to forego a systematic, focused and useful evaluation.

An evaluation is always shaped by the resources. Even when abundant or limitless resources are available there is a need for a focusing of it.

By having some guidelines to use, a decision-maker can be aware of the shortcuts and shortcomings of an evaluation as well as the strong points and advantages of an evaluation. Because there are limited resources does not mean that the decision-maker should reject evaluation. In the final instance, evaluation, or providing data for decision-makers, is meant to help the decision-maker, not hinder him. The suggestions provided herein are intended to aid the decision-maker in the evaluative process.

XIII. A GLOSSARY OF TERMS

Behavioral Objective: a statement of what you want someone (usually a learner) to accomplish, stated in very specific, behavioral terms.

Data for decision making: This is the statement of the purpose of educational evaluation, first set forth by Cronbach in 1963 and now widely held by the leading experts in the field, including Stufflebeam, Hutchinson, Guba, Alkin, Provus and so on.

Decision Maker: Any person who in some way makes a decision about a particular project, program, endeavor or enterprise. For a school, examples would be: students, parents, teachers, administrators, staff, school committee, etc.

Enterprise: That about which data is to be collected, that which is to be evaluated; can range from a single lecture to a whole program or project (e.g., Title I or III), to a school, to a national program.

Evaluation: the act of identifying, collecting, and reporting data to decision makers for their decision making needs.

Fuzzy Concept: Anything which is not directly observable or measurable is a fuzzy concept; a goal which is nebulous, vague, general, e.g., good citizen, autonomous learner, self-actualization.

Goal: A statement of intent or an aspiration, something you want to accomplish; usually stated in fuzzy terms.

Methodology: A standardized, operationalized, systematic set of rules and procedures for accomplishing a defined purpose.

Model: A generalized, non-specific set of general rules-of-thumb or guidelines for accomplishing a purpose; a set of non-operational, fuzzy procedures for doing something.

Observational Technique: Something with which to collect data, not just limited to a "test".

Operationalize: To take a fuzzy concept and systematically put it into its specific, concrete, observable, measurable state.

XIII. A GLOSSARY OF TERMS
(cont'd)

Prioritize: To put in some kind of order, e.g., putting a list of items in order of most important to least important or from first occurring in time to last occurring in time.

Resources: A term referring to money, time, staff, materials, space, expertise: those things which are needed to carry out an evaluation.

XIV. REFERENCES

1. Atkin, M. "Some Evaluation Problems in a Course Content Improvement Project", J Res Sci Ed, Vol. 1, (1963), pp. 129-132.

2. Ausabel, D.P. "Crucial Psychological Issues in the Objectives, Organization, and Evaluation of Curriculum Reform Movements", Psychology in the Schools, Vol. IV, No. 2, (April 1967), pp. 111-121.

3. Bloom, B. (Ed.) Taxonomy of Educational Objectives I: Cognitive Domain, New York: Longmans, Green, 1956.

4. Cronbach, L.J. Evaluation for course improvement. Teachers College Record, 1963, 64, pp. 231-248.

5. Eisner, E. "Instructional and Expressive Objectives: Their Formulation and Use in Curriculum", in Instructional Objectives, AERA Monograph Series, No. 3. Chicago: Rand McNally, 1969.

6. Guba, E.G., and Stufflebeam, D.L. Evaluation: the process of stimulating, aiding, and abetting insightful action. Address delivered at the Second National Symposium for Professors of Educational Research, Phi Delta Kappa, Boulder, Colorado, 1968.

7. Hutchinson, T.E. "A Numerical Example of Centour Analysis among Flexibly Determined Subgroups", American Educational Research Journal, Vol. 6, No. 1, 1969.

8. Hutchinson, T.E. Level of Aspiration and Statistical Models Applicable to the Problem of Refining Choice Bases for Career Development: Logic with Implications, (unpublished doctoral dissertation), (Xerox) Harvard Graduate School of Education, 1969.

9. Hutchinson, T.E., and Benedict, L.G. The operationalization of fuzzy concepts. University of Massachusetts. Xerox, 1970.

10. Mager, R.F. Preparing Instructional Objectives. Palo Alto: Fearon Publishers, 1962.

11. National Study of Secondary School Evaluation: Evaluative Criteria. Washington, D.C.: The Society, 1960.

11A. Pace, C.R. Evaluation perspectives: '68. Paper presented at an AERA Pre-session, Chicago, Illinois, 1968.

12. Popham, W.J. "Probing the Validity of Arguments Against Behavioral Goals", A Symposium Presented at AERA, Chicago, February 1968.

XIV. REFERENCES (cont'd)

13. Popham, W.J. 'Objectives and Instruction , in Instructional Objectives, AERA Monograph Series, No. 3, Chicago: Rand McNally, 1969.

14. Popham, W.J. and Baker, E. Establishing Instructional Goals. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

15. Raths, J.D. "Specificity as a Threat to Curriculum Reform", Paper presented at the AERA meetings, Chicago, February, 1968.

16. Scriven, M. The methodology of evaluation. In R.W. Tyler (Ed.), Perspectives of curriculum evaluation, AERA Curriculum Evaluation Monograph Series, #1. Chicago: Rand McNally, 1967.

17. Stake, R.E. The countenance of educational evaluation. Teachers College Record, 1967a, 68 (7), pp. 523-540.

18. Stake, R.E., Denny, T. "Needed Concepts and Techniques for Utilizing More Fully the Potential of Evaluation", in Educational Evaluation: New Roles, New Means, NSSE Yearbook, Part II, 1969, pp. 370-390.

19. Stufflebeam, D.L. A depth study of the evaluation requirement. Theory into Practice, 1967a, 5 (3), pp. 121-133.

20. Stufflebeam, D.L. The use and abuse of evaluation in title III. Theory into Practice, 1967b, 6 (5), pp. 126-133.

21. Stufflebeam, D.L. Evaluation as enlightenment for decision making. In W. Beatty (Ed.), Improving educational assessment. Washington, D.C.: Association for Supervision and Curriculum Development, 1969.

XV. ADDITIONAL REFERENCES

Benedict, L.G. A Survey of Goals. Center for Educational Research, University of Massachusetts. Xerox, 1970.

Benedict, L.G., and McKay, K. Program evaluation of the Mark's Meadow early childhood program: progress report, #1. Prepared and submitted to the Bureau of Curriculum Innovation, Massachusetts State Department of Education, Boston, November, 1970.

Benedict, L.G., and McKay, K. Program evaluation of the Mark's Meadow early childhood program: final report for the year 1970-71. Prepared and submitted to the Bureau of Curriculum Innovation, Massachusetts State Department of Education, Boston, June, 1971.

Coffing, R.T., Hutchinson, T.E., Thomann, J.B., and Allen, R.G. Self-instructional module for learning the Hutchinson method for operationalizing a goal or intent. Center for Educational Research, University of Massachusetts. Xerox, 1971

EPIC Evaluation Center, EPIC Brief #2, Tucson, Arizona, undated.

Gagne, R. Curriculum research and the promotion of learning. Perspectives of Curriculum Evaluation, AERA Monograph Series, #1. 1967, pp. 19-38.

Glass, G.V. The growth of evaluation methodology. University of Colorado, mimeo, March, 1969.

Gordon, G.M. Empirical testing of an evaluation methodology--the negotiation of the contract. A paper presented at the Graduate Colloquium, School of Education, University of Massachusetts, April, 1972.

Gordon, G.M. A field test of the Fortune/Hutchinson Evaluation methodology as it could be employed in the evaluation of national urban league street academies. Unpublished doctoral dissertation, University of Massachusetts, 1973.

Guba, E. Significant differences. Educational Researcher, XX:3, 1969, pp. 4-5.

Harris, C.W. Some issues in evaluation. The Speech Teacher, 1963, 12, pp. 191-199.

Hastings, J.T. Curriculum evaluation: the why of the outcomes. Journal of Educational Measurement, 3:1, 1966, pp. 27-32.

XV. ADDITIONAL REFERENCES (cont'd)

Hodson, W.A., and Watts, H. The first chance evaluation report for 1970-71. First Chance, Pre-School Education Centers for Brattleboro and Townshend, Brattleboro, Vermont, June, 1971.

Hutchinson, T.E. Some overlooked implications of the purpose: to provide data for decision making. A paper presented at AERA, Chicago, 1972.

Jones, L. The operationalization of educational objectives for the evaluation of an on-going program. Unpublished doctoral dissertation, University of Massachusetts, 1970.

Kresh, E. An overview of the discrepancy evaluation model and a related case study. Office of Research, Pittsburgh Public Schools. Mimeo, 1969.

Provus, M. Evaluation of on-going programs in the public school system. In R.W. Tyler (Ed.), Educational evaluation: new roles, new means II. Chicago: National Society for the Study of Education, 1969.

Provus, M. Discrepancy evaluation for educational program improvement and assessment. Berkeley, California: McCutchan, 1971.

Scriven, M.S. An introduction to meta-evaluation. Educational Product Report, 1969, 2 (5), pp. 36-38.

Scriven, M.S. Goal-free evaluation. Unpublished manuscript, University of California at Berkeley, 1971.

Stake, R. Toward a technology for the evaluation of educational programs. In R.W. Tyler (Ed.), Perspectives of Curriculum evaluation, AERA Curriculum Evaluation Monograph Series, #1. Chicago: Rand McNally, 1967b.

Stake, R.E. Generalizability of program evaluation: the need for limits. Educational Products Report, February, 1969a.

Stake, R.E. Language, rationality and assessment. In W.H. Beatty (Ed.), Improving Educational Assessment. Washington, D.C.: Association for Supervision and Curriculum Development, 1969b.

XV. ADDITIONAL REFERENCES (cont'd)

Stufflebeam, D.L., Foley, W.J., Gephart, W.J., Guba, E.G., Hammond, R.I., Merriman, H.O., and Provus, M.M. Educational evaluation and decision making. Itasca, Illinois: F.E. Peacock, 1971.

Wiley, D.E. Design and analysis of evaluation studies. In M.C. Wittrock and D.E. Wiley (Eds.), The evaluation of instruction: issues and problems. New York: Holt, Rinehart and Winston, 1970.